

## SUMMARIES

RESULT	1
LOCUS	AC092873
DEFINITION	Pan troglodytes clone RP43-22112, WORKING DRAFT SEQUENCE, 5
ACCESSION	AC092873
VERSION	AC092873.1
KEYWORDS	GT; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Chimpanzee.
ORGANISM	Pan troglodytes
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
	1 (bases 1 to 171413)
	Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
	Bouffard, G.G., Bren, K., Brinkley, C., Brooks, S., Dietrich, N.L.,
	Grahn, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,
	Lee-Hin, S.-O., Legaspi, R., Lim, M., Maduro, Q.L., Maduro, V.B.,
	Mastaglio, C., Mastrian, S.D., McCloskey, J.C., McDowell, U.,
	Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stantiford, S.,
	Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsurgren, C., Vogt, J.L.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT

Walker, M.A., Wetherby, K.D., Zhang, L.-H. and Green, E.D.  
NISC Comparative Sequencing Initiative  
Unpublished  
2 (bases 1 to 171413)  
Green, E.D.  
Direct Submission  
Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717  
Government Circle, Gaithersburg, MD 20877, USA  
Genome Center  
Center: NIH Intramural Sequencing Center  
Center code: NISC  
Web site: <http://www.nisc.nih.gov>  
Contact: [nisc.mouse@nih.gov](mailto:nisc.mouse@nih.gov)  
Project Information  
Center project name: ani  
Center clone name: 022112  
Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 170445 bases at least Q40  
Consensus quality: 170634 bases at least Q20  
Insert size: 152000; agarose-gel  
Insert size: 168000; pulse-field-gel  
Quality coverage: 15.61x in Q20 bases; agarose-gel  
Quality coverage: 14.12x in Q20 bases; pulse-field-gel  
Quality coverage: 13.87x in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 2901: contig of 2901 bp in length  
\* 2902 3001: gap of unknown length  
\* 3002 7859: contig of 4858 bp in length  
\* 7860 7959: gap of unknown length  
\* 7960 17459: contig of 9500 bp in length  
\* 17460 17559: gap of unknown length  
\* 17560 90578: contig of 73019 bp in length  
\* 90579 171413: gap of unknown length  
\* 171413: contig of 80735 bp in length.  
Location/Qualifiers  
1. 171413  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="RP43-22112"  
/clone\_11b="RP43"  
1. 2901  
/note="assembly\_fragment"  
clone\_end:77  
vector\_side:left"  
3002.7859  
/note="assembly\_fragment"  
7960.17459  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:right"  
17560.90578  
/note="assembly\_fragment"  
90679.171413  
/note="assembly\_fragment"  
BASE COUNT 53796 a 33490 c 32103 g 51618 t 406 others  
ORIGIN

Query Match 18.6%; Score 35.8; DB 2; Length 171413;  
Best Local Similarity 63.2%; Pred. No. 1.8;

Matches 55: Conservative 0: Mismatches 32: Indels 0: Gaps 0:

QY 15 ccataaagtcgtacacgaatactctgtctgattcgtcatttcataaccctcgagg 74  
DB 113527 CCAGCAGCAGGCGGTACATAGTATCATATTTTAAATTCCTCATGTATTAACACATTTTAT 113468  
QY 75 gaacatagatgtaaaccttataaa 101  
DB 113467 ATRAGCATTTTAAAAACACTTAAAAA 113441

RESULT 2  
AC087432  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-9664 map 3p, complete sequence.  
ACCESSION AC087432.2 GI:12745086  
VERSION AC087432  
KEYWORDS  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 152423)  
Xiong, H., Zhou, Y., Dong, H., Lin, W., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Yang, H.F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z. and Huang, M.  
Unpublished  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 152423)  
Lin, B., Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, Y., Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.  
Direct Submission  
Submitted (05-JAN-2001) 1. Chinese Human Genome Center at Shanghai  
2. Institute of Genetics, Chinese Academy of Sciences, P.R.China  
3 (bases 1 to 152423)  
Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.  
Direct Submission  
Submitted (11-FEB-2001) Human Genome Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China  
On Feb 11, 2001 this sequence version replaced gi.12039244.  
Genome Center  
Center: Beijing Center  
Center code: Beijing  
Website: <http://hgsc.igtp.ac.cn>  
<http://www.genomics.org.cn>  
Contact: [hgsc@igtp.ac.cn](mailto:hgsc@igtp.ac.cn)  
Project Information  
Center project name: 11 project  
Center clone name: RP11-9664  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator; ET 55% of reads

TITLE  
JOURNAL  
COMMENT

```
Chemistry: Dye-terminator Big Dye; 458 of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 890 bases at least Q40
Consensus quality: 1126 bases at least Q30
Consensus quality: 1224 bases at least Q20
Insert size: 1206; sum-of-coverage
Quality coverage: 1.11x in Q20 bases; sum-of-coverage
-----
FEATURES
  source
    Location/Qualifiers
      1..152423
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="3"
        /map="3p"
        /clone="RP11-96G4"
BASE COUNT      50230 a 28252 c 28134 g 45807 t
ORIGIN
Query Match      18.1%; Score 34.8; DB 9; Length 152423;
Best Local Similarity 65.4%; Pred. No. 3.8;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 47 tgattgctattttcaaaccttcggggaatagatgtgaacccttataaacgcgg 106
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6179 TGAAGAAGACTATTGTACACCTTGTGCAGAAATATTTGAAAATCTAGATGAATACAA 6238
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 107 gtttcgcagaacatgc 124
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6239 TATTTCCTTAGAAAAATAC 6256
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
AC020753 160410 bp DNA HTG 03-FEB-2000
LOCUS Homo sapiens chromosome 3p clone RP11-322F9, WORKING DRAFT
DEFINITION SEQUENCE, 14 unordered pieces.
ACCESSION AC020753
VERSION AC020753.2 GI:6862663
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 160410)
Wang, L., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H.,
Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H.,
Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Song, L., Zhang, L.,
Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
Wang, X., Feng, X., Yu, J. and Yang, H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 160410)
Wang, L., Wang, X., Hu, S., Dong, W., Wang, J., Zhang, X., Zhang, Y.,
Zhang, H., Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X.,
Cheng, C., Wang, Y., Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wu, D.,
Yu, B., Fan, H., Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Song, L.,
Song, L., Zhang, L., Guo, D., Huang, F., Zhang, G., Li, J., Bian, X.,
Zhang, M., Li, L., Feng, X., Yu, J. and Yang, H.
Direct Submission
Submitted (09-JAN-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P. R. China
On Feb 3, 2000 this sequence version replaced gi:6864192.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
1 2071: contig of 2071 bp in length
* gap of unknown length
* 2072 6327: contig of 4256 bp in length
* gap of unknown length
* 6328 9716: contig of 3389 bp in length
* gap of unknown length
* 9717 13515: contig of 3799 bp in length
* gap of unknown length
* 13516 23667: contig of 10152 bp in length
* gap of unknown length
* 23668 32423: contig of 8756 bp in length
* gap of unknown length
* 32424 43384: contig of 10961 bp in length
* gap of unknown length
* 43385 53196: contig of 9812 bp in length
* gap of unknown length
* 53197 62912: contig of 9716 bp in length
* gap of unknown length
* 62913 73306: contig of 10394 bp in length
* gap of unknown length
* 73307 89106: contig of 15800 bp in length
* gap of unknown length
* 89107 108897: contig of 19791 bp in length
* gap of unknown length
* 108898 133544: contig of 24647 bp in length
* gap of unknown length
* 133545 160410: contig of 26866 bp in length.
FEATURES
  source
    Location/Qualifiers
      1..160410
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="3p"
        /clone="RP11-322F9"
BASE COUNT      52178 a 30289 c 29805 g 48138 t
ORIGIN
```

```
Query Match      18.1%; Score 34.8; DB 2; Length 160410;
Best Local Similarity 65.4%; Pred. No. 3.8;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 47 tgattgctattttcaaaccttcggggaatagatgtgaacccttataaacgcgg 106
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139713 TGAAGAAGACTATTGTACACCTTGTGCAGAAATATTTGAAAATCTAGATGAATACAA 139772
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 107 gtttcgcagaacatgc 124
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139773 TATTTCCTTAGAAAAATAC 139790
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
AC018814 166156 bp DNA PRI 06-MAR-2001
LOCUS Homo sapiens chromosome 3 clone RP11-204C23 map 3p, complete
DEFINITION sequence.
ACCESSION AC018814
VERSION AC018814.5 GI:13236630
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 166156)
Xiong, H., Zhou, Y., Dong, H., Liu, W., Chen, B., Zhang, C., Zhang, Y.,
Cai, Z., Ying, H., F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J.,
Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G.,
Chen, Z., and Huang, M.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 166156)
Li, G., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W.,
Sun, Y., Wu, Q., Dong, W., Wang, J., Zhang, Y., Yang, X., Niu, Y., Qi, X.,
Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Wang, X.,
```

	TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
REFERENCE	3 (bases 1 to 166156)	
AUTHORS	Zhou,Y., Xiong,H., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y., Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J., Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G., Chen,Z. and Huang,M.	
TITLE	Direct Submission	
JOURNAL	Submitted (27-DEC-2000) 1. Chinese Human Genome Center at Shanghai	
REFERENCE	2. Institute of Genetics, Chinese Academy of Sciences, P.R.China	
AUTHORS	4 (bases 1 to 166156)	
	Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xian,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,W., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J. and Yang,H.	
TITLE	Direct Submission	
JOURNAL	Submitted (06-MAR-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China	
COMMENT	On Mar 6, 2001 this sequence version replaced gi:11992941. -----Genome Center Center:Beijing Center Center code:Beijing Website:http://hgsc.tgtp.ac.cn http://www.genomics.org.cn Contact:ngc@tgtp.ac.cn Project Information Center project name:1% project Center clone name: RP11-204C23 -----Summary Statistics Sequencing vector: pUC18; 100% of reads Chemistry: Dye-terminator; ET 5% of reads Assembly program: Phrap; Version 0.990329 Consensus quality: 314 bases at least Q40 Consensus quality: 468 bases at least Q30 Consensus quality: 518 bases at least Q20 Insert size: 534; sum-of-contigs Quality coverage: 4.75x In Q20 bases,sum-of-contigs ----- Location/Qualifiers 1. 166156 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" /map="3p" /clone="RP11-204C23"	
BASE COUNT	55129 a 31315 c 30797 g 48915 t	
ORIGIN		
Query Match	18.1%; Score 34.8; DB 9; Length 166156;	
Best Local Similarity	65.4%; Pred. No.3.8;	
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;		
47	tgtattcgcttttccatcaacttcggggaataagatgtgaaacocctataaacgcg 106	
Dn 108901	TGAAAGACGCTATTTTGTCACCGCTTGTGCACAATATATTTTGAAAATCTAGATGAATAACAA 108960	
Oy	107 gtcttcgcagaacaatgc 124	

Db	108961	TATTTCTTAGAAATAC	108978	
RESULT	5			
AC087085				
ID	AC087085	standard; DNA; HUM; 166156 BP.		
AC	AC087085;			
XX	SV	AC087085.1		
XX				
DT	07-DEC-2000	(Rel. 66, Created)		
DT	07-DEC-2000	(Rel. 66, last updated, Version 1)		
XX				
DE	Homo sapiens chromosome 3p clone RP11-204c23, complete sequence.			
XX				
XX	HTG.			
XX				
OS	Homo sapiens (human)			
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia			
CC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
XX				
XX	[1]			
RP	1-166156			
RA	Zhou X., Xiong H., Dong H., Lin W., Chen B., Zhang C., Zhang Y., Cai Z.,			
RA	Ying H.F., Wang H., Gu W., Zhu G., Tu Y., Zhang X., Jia J., Shen H.,			
RA	Zhang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,			
RA	Huang M.;			
RT	"Chromosome 3p genomic sequence";			
RL	unpublished.			
XX				
RN	[2]			
RP	1-166156			
RA	Zhou X., Xiong H., Dong H., Lin W., Chen B., Zhang C., Zhang Y., Cai Z.,			
RA	Ying H.F., Wang H., Gu W., Zhu G., Tu Y., Zhang X., Jia J., Shen H.,			
RA	Zhang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,			
RA	Huang M.;			
RT	Submitted (06-DEC-2000) to the EMBL/GenBank/DBJ databases.			
RL	1. Chinese Human Genome Center at Shanghai 2. Institute of Genetics,			
RL	Chinese Academy of Sciences, P.R.China			
XX				
CC	-----Genome Center-----			
CC	Center:Beijing Center Center code:Beijing			
CC	Website:http://www.chgc.sh.cn http://hgc.igtp.ac.cn			
CC	http://www.genomics.org.cn Contact:hgc@igtp.			
CC	NOTE: This record was originally submitted by Chinese Human Genome			
CC	Center at Shanghai (CHGC), and now maintained by CHGC, submitted by			
CC	Beijing Center.			
XX				
FH	Key	Location/Qualifiers		
FT	source	1. 166156		
FT		/chromosome="3p"		
FT		/db_xref="taxon:9606"		
FT		/organism="Homo sapiens"		
FT		/clone="RP11-82c3"		
FT		/clone="RP11-204c23"		
FT		1030. 1056		
FT		/rpt_family="Simple_repeat"		
FT		/rpt_unit="(GAA)n		
FT		1250. 1371		
FT		/rpt_family="SINE/MIR"		
FT		/rpt_unit=MIR		
FT		1647. 1989		
FT		/rpt_family="LINE/L1"		
FT		/rpt_unit=L1M4		
FT		2019. 2158		
FT		/rpt_family="LTR/Retroviral"		
FT		/rpt_unit=MLT2FB		
FT		2167. 2705		
FT		/rpt_family="LINE/L1"		
FT		/rpt_unit=L1MC3		



```
FT repeat_region 2709 .3765 /rpt_family="LINE/L1"
FT /rpt_unit-L1MC3
FT repeat_region 3766 .3803 /rpt_unit-L1MC3
FT /rpt_family="Simple_repeat"
FT /rpt_unit-(CA)n
FT repeat_region 3804 .4627 /rpt_family="LINE/L1"
FT /rpt_unit-L1MC3
FT repeat_region 4664 .4791 /rpt_family="LINE/L1"
FT /rpt_unit-L1MC3
FT repeat_region 4819 .4980 /rpt_family="SINE/MIR"
FT /rpt_unit-MIR
FT repeat_region complement(5935 .6051)
FT /rpt_family="LINE/L2"
FT /rpt_unit-L2
FT repeat_region complement(6792 .9493)
FT /rpt_family="LINE/L1"
FT /rpt_unit-L1PA5
FT repeat_region complement(9544 .9591)
FT /rpt_family="DNA/Mariner"
FT /rpt_unit-MADE1
FT repeat_region 10733 .10885 /rpt_family="DNA/MER1_type"
FT /rpt_unit-MER63A
FT repeat_region complement(13506 .13551)
FT /rpt_family="SINE/MIR"
FT /rpt_unit-MIR
FT repeat_region 14554 .14587 /rpt_family="Low_complexity"
FT /rpt_unit-AT-rich
FT repeat_region complement(14646 .14840)
FT /rpt_family="LINE/L1"
FT /rpt_unit-L1M4
FT repeat_region 14866 .15024 /rpt_family="DNA/MER1_type"
FT /rpt_unit-MER33
FT repeat_region complement(15025 .15166)
FT /rpt_family="LINE/L1"
FT /rpt_unit-L1MCB
FT repeat_region complement(15167 .15478)
FT /rpt_family="SINE/Alu"
FT /rpt_unit-AluSg
FT repeat_region complement(15479 .15491)
FT /rpt_family="LINE/L1"
FT /rpt_unit-L1MCB
FT repeat_region 16180 .16238 /rpt_family="LTR/MaLR"
FT /rpt_unit-MTR1E2
FT repeat_region 16288 .16642 /rpt_family="LTR/MaLR"
FT /rpt_family="LTR/MaLR"
FT repeat_region 16666 .17129 /rpt_family="LTR/MaLR"
FT /rpt_unit-MTR1E2
FT repeat_region 17198 .17236 /rpt_family="Low_complexity"
FT /rpt_unit-AT-rich
FT repeat_region complement(18160 .18573)
FT /rpt_family="LINE/L2"
FT /rpt_unit-L2
FT STS 18303 .18735 /standard_name="SHGC-82788"
FT 19992 .20045 /rpt_family="SINE/MIR"
FT /rpt_unit-MIR
FT repeat_region complement(20515 .20823)
FT /rpt_family="SINE/Alu"
FT /rpt_unit-AluSg
FT repeat_region complement(20835 .21083)
FT /rpt_family="LINE/L1"

FT repeat_region 21753 .22022 /rpt_family="SINE/Alu"
FT /rpt_unit-AluSg
FT repeat_region 22023 .22053 /rpt_family="Low_complexity"
FT /rpt_unit-AT-rich
FT repeat_region 22196 .22643 /rpt_family="LINE/L1"
FT /rpt_unit-L1PA8
FT repeat_region complement(23919 .23963)
FT /rpt_family="LINE/L2"
FT /rpt_unit-L2
FT repeat_region 24040 .24249 /rpt_family="SINE/MIR"
FT /rpt_unit-MIR
FT repeat_region 24302 .24801 /rpt_family="SINE/Alu"
FT /rpt_unit-AluSx
FT repeat_region complement(24859 .24984)
FT /rpt_family="LINE/L2"
FT /rpt_unit-L2
FT repeat_region 25258 .25618 /rpt_family="LTR/MaLR"
FT /rpt_unit-THE1B
FT repeat_region 25849 .26929 /rpt_family="LINE/L1"
FT /rpt_unit-L1PA12
FT repeat_region 26974 .27007 /rpt_family="Low_complexity"
FT /rpt_unit-AT-rich
FT repeat_region 27907 .28148 /rpt_family="SINE/Alu"
FT /rpt_unit-AluSx
FT repeat_region 28152 .28179 /rpt_family="Low_complexity"
FT /rpt_unit-AT-rich
FT repeat_region 28669 .28779 /rpt_family="LINE/L1"
FT /rpt_unit-L1MEC
FT repeat_region 29051 .29290 /rpt_family="LINE/L1"
FT /rpt_unit-L1MEC
FT repeat_region 29342 .30136 /rpt_family="LINE/L1"
FT /rpt_family="Simple_repeat"
FT repeat_region 30160 .30331 /rpt_unit-(TA)n
FT repeat_region 30339 .30487 /rpt_family="LINE/L1"
FT /rpt_unit-L1MEC
FT repeat_region 30532 .30819 /rpt_family="SINE/Alu"
FT /rpt_unit-AluSx
FT repeat_region 30901 .31101 /rpt_family="LINE/L1"
FT /rpt_unit-L1M4
FT repeat_region 31102 .31410

Query Match 18.1%; Score 34.8; DB 17; Length 166156;
Best Local Similarity 65.4%; Pred.No.3.8;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 47 tgcctcgtctatttcacacttcctggggaatagatgtgaacccttataaagcgcg 106
|||||
Db 108901 TGAAGAAGCTATTGTTGTACACCTTGTCGAATATATTGAAATCTAGATGAATACAA 108960
|||||
Qy 107 gtttcgcagaacatgc 124
|||||
Db 108961 TATTCTTAGAAAAATAC 108978
```

RESULT 6  
AC024216/c LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-204C23, WORKING DRAFT  
ACCESSION AC024216 AC006431  
SEQUENCE 35 unordered pieces.  
AC024216.16 GI:9438304  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 181870)  
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,  
Bodola,B., Bouck,J., Bowie,S., Brooks,A., Bunay,C., Bunac,C.,  
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,  
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,  
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,  
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,  
Guevara,M., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,  
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,  
Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,  
Lichtarge,O., Liu,J., Liu,M., Logan,O., Lozano,R.J., Lu,J.,  
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M.,  
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,  
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,  
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,  
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,S.,  
Taber,P., Taylor,T., Vasquez,L., Vanson,R., Vo,O., Wahbah,M.,  
Wallington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,  
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and  
GIBBS,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 181870)  
Worley,K.C.  
Direct Submission  
Submitted (28-FEB-2000) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 25, 2000 this sequence version replaced g1:9255957.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HAGO  
Center clone name: RP11-204C23  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 155457 bases at least Q40  
Consensus quality: 168744 bases at least Q30  
Consensus quality: 176318 bases at least Q20  
Estimated insert size: 168203; sum-of-coverage estimation  
Quality coverage: 0x in Q20 bases; average-tp estimation  
Quality coverage: 3.3x in Q20 bases; sum-of-coverage estimation  
----- NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 35 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 29896: contig of 29896 bp in length  
\* 29897 29996: gap of unknown length  
\* 29997 44416: contig of 14420 bp in length  
\* 44417 44516: gap of unknown length  
\* 44517 63362: contig of 18846 bp in length

63363 63462: gap of unknown length  
\* 63463 78902: contig of 15440 bp in length  
\* 78903 79002: gap of unknown length  
\* 79003 89873: contig of 10871 bp in length  
\* 89874 97597: gap of unknown length  
\* 97597 97697: contig of 7624 bp in length  
\* 97698 105158: gap of unknown length  
\* 105159 105258: gap of unknown length  
\* 105259 112486: contig of 7228 bp in length  
\* 112487 112587: gap of unknown length  
\* 112587 117617: contig of 5031 bp in length  
\* 117618 117717: gap of unknown length  
\* 117718 123759: contig of 6042 bp in length  
\* 123760 123859: gap of unknown length  
\* 123860 128429: contig of 4570 bp in length  
\* 128430 128529: gap of unknown length  
\* 128530 132354: contig of 3825 bp in length  
\* 132355 132454: gap of unknown length  
\* 132455 135891: contig of 3437 bp in length  
\* 135892 135991: gap of unknown length  
\* 135992 141318: contig of 5327 bp in length  
\* 141319 141418: gap of unknown length  
\* 141419 145280: contig of 3862 bp in length  
\* 145281 145380: gap of unknown length  
\* 145381 148462: contig of 3082 bp in length  
\* 148463 148562: gap of unknown length  
\* 148563 151476: contig of 2914 bp in length  
\* 151477 151576: gap of unknown length  
\* 151577 154759: contig of 3183 bp in length  
\* 154760 154859: gap of unknown length  
\* 154860 156425: gap of 1566 bp in length  
\* 156426 156525: gap of unknown length  
\* 156526 158715: contig of 2190 bp in length  
\* 158716 158815: gap of unknown length  
\* 158816 161739: contig of 2924 bp in length  
\* 161740 161839: gap of unknown length  
\* 161840 163733: contig of 1534 bp in length  
\* 163734 163473: gap of unknown length  
\* 163474 164495: contig of 1022 bp in length  
\* 164496 164595: gap of unknown length  
\* 164596 166256: contig of 1661 bp in length  
\* 166257 166356: gap of unknown length  
\* 166357 168039: contig of 1683 bp in length  
\* 168040 168139: gap of unknown length  
\* 168140 169184: contig of 1045 bp in length  
\* 169185 169284: gap of unknown length  
\* 169285 170463: contig of 1179 bp in length  
\* 170464 170563: gap of unknown length  
\* 170564 172541: contig of 1978 bp in length  
\* 172542 172641: gap of unknown length  
\* 172642 174280: contig of 1633 bp in length  
\* 174281 174380: gap of unknown length  
\* 174381 175663: contig of 1283 bp in length  
\* 175664 175763: gap of unknown length  
\* 175764 176977: gap of 1234 bp in length  
\* 176978 177097: gap of unknown length  
\* 177098 177098: contig of 1134 bp in length  
\* 177099 178331: gap of unknown length  
\* 178332 179583: contig of 1252 bp in length  
\* 179584 179683: gap of unknown length  
\* 179684 180695: contig of 1012 bp in length  
\* 180696 180795: gap of unknown length  
\* 180796 181870: contig of 1075 bp in length.  
Location/Qualifiers  
1. 181870  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-204C23"

BASE COUNT 53768 a 34063 c 34457 g 56115 t 3467 others  
ORIGIN

Query Match	Similarity	18.1%	Score 34.8	DB 2	Length 181870
Best Local	Similarity	65.4%	Pred No. 3.9		
Matches	51: Conservative	0:	Mismatches	27:	Indels
					Gaps 0:
Oy	47	tgattcgtcatttcacaccttcggyggaataagatgtaaacacctataaacgcgg	106		
Db	21069	TGAAAAGACATTTTGTACACCTTGTGCAAAATATTTGAAATCATGATGAATAACAA	21010		
Oy	107	gttttcgcggaacatgc	124		
Db	21009	TATTTCTTAGAAAAATAC	20992		
RESULT	7				
LOCUS	AC018497				
DEFINITION	Homo sapiens chromosome 3p clone RP11-229E21, WORKING DRAFT				
ACCESSION	AC018497				
VERSION	AC018497.4	GI:6862671			
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 188432)				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 188432)				
REFERENCE	Bao,Q., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,W., Sun,Y., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Li,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Wang,X., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J. and Yang,H.				
AUTHORS	Chromosome 3p genomic sequence				
TITLE	Submitted (13-DEC-1999)				
JOURNAL	Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China				
COMMENT	On Feb 3, 2000 this sequence version replaced gi:6864200.				
	* NOTE: This is a working draft sequence. It currently				
	* consists of 11 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is				
	* arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				
	* This record will be updated with the finished sequence				
	* as soon as it is available and the accession number will				
	* be preserved.				
	1	2725: contig of 2725 bp in length			
	*	gap of unknown length			
	*	2726	4464: contig of 1739 bp in length		
	*	gap of unknown length			
	*	4465	5842: contig of 1378 bp in length		
	*	gap of unknown length			
	*	5843	9661: contig of 3819 bp in length		
	*	gap of unknown length			
	*	9662	16331: contig of 6670 bp in length		
	*	gap of unknown length			
	*	16332	24690: contig of 8359 bp in length		
	*	gap of unknown length			
	*	24691	40720: contig of 16030 bp in length		
	*	gap of unknown length			
	*	40721	72584: contig of 31864 bp in length		
	*	gap of unknown length			
	*	72585	106317: contig of 31733 bp in length		
	*	gap of unknown length			

```

FEATURES
  SOURCE
    * 106318 142895: contig of 36578 bp in length
    * 142896 188432: contig of 45537 bp in length.
      Location/Qualifiers
        1..188432
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="3p"
          /clone="RP11-229E21"

BASE COUNT  6066 a 35322 c 35795 g 56644 t 3 others

ORIGIN
Query Match      18.1%; Score 34.8; DB 2; Length 188432;
Best Local Similarity 65.4%; Pred. No. 3.9;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0

OY 47 tgatcgcctatcttccataacctcggggaatagatgtgaaaaacctataaacgcgg 106
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133724 TGAAGAGCATATTTGTGTACACCTTTGTGCAATATATTTGAAATCATAGTGAATACAA 133783
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 107 gtcttcgcagaaacatgc 124
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133784 TATTCTTAGAAAAATAC 133801

RESULT 8
AC026486
LOCUS
DEFINITION Homo sapiens clone RP11-24K7, WORKING DRAFT SEQUENCE, 18 unordered
           pieces.
ACCESSION AC026486
VERSION AC026486.2 GI:9954792
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-24K7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
           Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
           Campolunghi,A., Castle,A., Chopel,Y., Colangelo,M., Collins,S.,
           Collamore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
           Klein,J., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
           Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
           McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
           Medirim,T., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
           Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisanik,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Testiere,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zalnoun,D., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           On Aug 31, 2000 this sequence version replaced gi:7280316.
           All repeats were identified using RepeatMasker:
           Smolt, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research

```

```

Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4489
Center clone name: 24_K-7
----- Summary Statistics
Sequencing vector: M13; M77813; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 140190 bases at least Q40
Consensus quality: 144274 bases at least Q30
Consensus quality: 145694 bases at least Q20
Insert size: 145000; agarose-ff
Insert size: 146569; sum-of-ctnigs
Quality coverage: 4.6 in Q20 bases; agarose-ff
Quality coverage: 4.5 in Q20 bases; sum-of-ctnigs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 18 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1. 1564: ctnig of 1564 bp in length
1565 1664: gap of 100 bp
1665 20700: ctnig of 1936 bp in length
20701 20800: gap of 100 bp
20801 21849: ctnig of 1049 bp in length
21850 21949: gap of 100 bp
21950 25604: ctnig of 3655 bp in length
25605 25704: gap of 100 bp
25705 29764: ctnig of 4060 bp in length
29765 29864: gap of 100 bp
29865 35579: ctnig of 5715 bp in length
35580 35679: gap of 100 bp
35680 41577: ctnig of 5898 bp in length
41578 41677: gap of 100 bp
41678 47413: ctnig of 5736 bp in length
47414 47513: gap of 100 bp
47514 54860: ctnig of 7347 bp in length
54861 54960: gap of 100 bp
54961 61569: ctnig of 6629 bp in length
61590 61689: gap of 100 bp
61690 68215: ctnig of 6526 bp in length
68216 68315: gap of 100 bp
68316 78737: ctnig of 10422 bp in length
78738 78837: gap of 100 bp
78838 89917: ctnig of 11080 bp in length
89918 90017: gap of 100 bp
90018 100805: ctnig of 10788 bp in length
100806 100905: gap of 100 bp
100906 114689: ctnig of 13784 bp in length
114690 114789: gap of 100 bp
114790 132632: ctnig of 17843 bp in length
132633 132732: gap of 100 bp
132733 147978: ctnig of 15246 bp in length
147979 148078: gap of 100 bp
148079 148269: ctnig of 191 bp in length.
Location/Qualifiers
1. 148269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-24K7"
/clone_1lb="RP11-11 Human Male BAC"
1. 1564
/note="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature
1665..20700
/note="assembly-fragment"

```



AUTHORS Wray, P.  
TITLE Direct Submission  
JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquay@sanger.ac.uk

## COMMENT

Requests: clonerequest@sanger.ac.uk  
On May 11, 2001 this sequence version replaced g1:11995192.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-63A2 is from the library RP11-11.1 constructed by the group of  
Pleier de Jong. For further details see  
<http://www.chori.org/Bacpac/home.htm>  
VECTOR: PBAC3.6  
This sequence is the entire insert of clone RP11-63A2. The true  
right end of clone RP11-809M12 is at 46921 in this sequence.

## FEATURES

source

Location/Qualifiers

1. 16973  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-63A2"  
/clone\_11b="RP11-11.1"  
3. 805  
/note="LMC5 repeat: matches 7127. .7913 of consensus"  
1250. 1679  
/note="MLT2B repeat: matches 1. .444 of consensus"  
1882. 4635  
/note="HERVL repeat: matches 2635. .5652 of consensus"  
4636. 4874  
/note="AluSq repeat: matches 67. .297 of consensus"  
4875. 7529  
/note="HERVL repeat: matches 1. .2635 of consensus"  
7530. 7577  
/note="MLT2A repeat: matches 406. .453 of consensus"  
7594. 8007  
/note="MLT2A repeat: matches 150. .450 of consensus"  
8333. 8517  
/note="MLT1-INTERNAL repeat: matches 932. .1211 of  
consensus"  
8555. 8854  
/note="LMC/D repeat: matches 5313. .5604 of consensus"  
8913. 9043  
/note="LMC/D repeat: matches 5604. .5735 of consensus"  
9108. 9312  
/note="LMC5 repeat: matches 7273. .7496 of consensus"  
9469. 9845  
/note="LMC5 repeat: matches 7507. .7895 of consensus"  
10014. 10105  
/note="46 copies 2 mer aa 63% conserved"  
10601. 10716  
/note="LM4 repeat: matches 2695. .2818 of consensus"  
10829. 11016  
/note="LM4 repeat: matches 2957. .3146 of consensus"  
11333. 11390

repeat\_region  
/note="MLT2D repeat: matches 1. .58 of consensus"  
11411. 11516  
/note="LM4 repeat: matches 2957. .3062 of consensus"  
11612. 11968  
repeat\_region  
/note="MLT2D repeat: matches 70. .435 of consensus"  
11986. 12049  
repeat\_region  
/note="32 copies 2 mer aa 67% conserved"  
12683. 12935  
repeat\_region  
/note="LM3A repeat: matches 5659. .5936 of consensus"  
13751. 14135  
/note="LME repeat: matches 471. .848 of consensus"  
14156. 14460  
repeat\_region  
/note="Aluub repeat: matches 1. .300 of consensus"  
14461. 14976  
repeat\_region  
/note="LME repeat: matches 848. .1572 of consensus"  
17233. 17352  
/note="3 copies 40 mer 75% conserved"  
17242. 17345  
repeat\_region  
/note="26 copies 4 mer gaag 72% conserved"  
17276. 17359  
repeat\_region  
/note="14 copies 6 mer agaag 65% conserved"  
18041. 18273  
repeat\_region  
/note="MIR repeat: matches 8. .246 of consensus"  
19550. 19633  
repeat\_region  
/note="42 copies 2 mer ta 83% conserved"  
19554. 19633  
repeat\_region  
/note="2 copies 40 mer 91% conserved"  
20024. 22103  
repeat\_region  
/note="LIPB3 repeat: matches 4126. .6150 of consensus"  
26100. 26159  
repeat\_region  
/note="35 copies 2 mer at 77% conserved"  
26101. 26164  
repeat\_region  
/note="16 copies 4 mer tata 79% conserved"  
26106. 26165  
repeat\_region  
/note="5 copies 12 mer 81% conserved"  
26204. 27088  
repeat\_region  
/note="LMC3 repeat: matches 6716. .7586 of consensus"  
27760. 27918  
repeat\_region  
/note="MLT1 repeat: matches 201. .346 of consensus"  
28811. 28944  
repeat\_region  
/note="MIR repeat: matches 12. .149 of consensus"  
29062. 29149  
repeat\_region  
/note="MLT1D repeat: matches 2. .89 of consensus"  
30829. 30884  
repeat\_region  
/note="MIR repeat: matches 76. .131 of consensus"  
33710. 33949  
/note="MIR repeat: matches 13. .262 of consensus"  
34018. 34106  
repeat\_region  
/note="12 repeat: matches 2655. .2747 of consensus"  
34360. 34452  
repeat\_region  
/note="MIR repeat: matches 50. .144 of consensus"  
35653. 35789  
repeat\_region  
/note="MIR repeat: matches 3. .137 of consensus"  
36130. 36201  
repeat\_region  
/note="12 copies 6 mer atatat 69% conserved"  
36131. 36202  
repeat\_region  
/note="36 copies 2 mer ta 70% conserved"  
37945. 37968  
repeat\_region  
/note="12 copies 2 mer tc 100% conserved"  
38866. 39253  
repeat\_region  
/note="LM3 repeat: matches 5912. .6304 of consensus"  
39256. 39337  
repeat\_region  
/note="THEIR repeat: matches 1. .82 of consensus"  
39340. 41088  
repeat\_region  
/note="LM1 repeat: matches 971. .2727 of consensus"  
41115. 41324  
repeat\_region  
/note="LMC5 repeat: matches 7701. .7912 of consensus"  
41359. 41767  
repeat\_region  
/note="LM8 repeat: matches 5725. .6145 of consensus"  
41771. 41943  
repeat\_region  
/note="LMC5 repeat: matches 7603. .7774 of consensus"  
43197. 43436  
repeat\_region  
/note="MIR repeat: matches 16. .260 of consensus"

[illegible]

REFERENCE 1 (bases 1 to 188026)  
AUTHORS Blakey, S.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire  
C910 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk  
requests: clonerequests@sanger.ac.uk  
COMMENT On JUL 25, 2001 this sequence version replaced gi:14970342.

**FEATURES**

- \* **NOTE:** This is a 'working draft' sequence.
- \* This sequence will be replaced
- \* by the finished sequence as soon as it is available and
- \* the accession number will be preserved.

**Location/Qualifiers**

```
misc_feature      61435 a 34467 c 34626 g 57498 t
BASE COUNT
ORIGIN
```

Query Match	Best Local Similarity	17.9% 52.0%	Score 34.4	DB 2	Length 188026				
Matches	77	Conservative	0	Mismatches	71	Indels	0	Gaps	0
QY	27	ttgcaagcataactctgctgcattcgctcgtatatttcaataacctcggggaaatagatg	86						
Db	179899	TTTACCAATTATATGTTTCATTGATATGTTGATGTTCAAGACATTTGCTCTCATATATTGG	179840						
QY	87	aaaacctataaacaacgcgggttttcgcagaaacaatgcgctagatcattgatgaaca	146						
Db	179839	ATTTACCAATTAAGATATCAAAATTTGTCATGCGCATTTATTTATATAGATCATTTGA	179780						
QY	147	tggactaagcaaaagtcgtcctctg	174						
Db	179779	TGTACTAGAGGAAAATTTCTCTCTCTTG	179752						

RESULT 12	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
VIYD10/c	VIYD10	155711 bp DNA	INV	06-JUL-2000					
		Plasmodium vivax telomeric YAC clone, complete finished sequence.							
		AL360354							
		AL360354.1	GI:8953641						
						malaria parasite P. vivax.			
						Plasmodium vivax			
						Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
						1 (bases 1 to 155711)			
						Oliver,K., Bowman,S., Hall,N., Quall,M., Rajandream,M.A.,			

Harris, D., del Portillo, H.A., Lanzer, M. and Barrell, B.G.  
Direct Submission  
Submitted (27-JUN-2000) The Sanger Centre, Wellcome Trust Genome  
Campus, Hinxton, Cambridge CB10 1SA, UK  
For more information about this project  
see [http://www.sanger.ac.uk/Projects/P\\_vivax](http://www.sanger.ac.uk/Projects/P_vivax).  
Location/Qualifiers  
1. 155711  
/organism="Plasmodium vivax"  
/db\_xref="taxon:5855"  
/country="Brazil"  
1. 611  
complement(join(1173. .1337,1442. .2149,2276. .2302))  
/gene="v1r1"  
complement(join(1173. .1337,1442. .2149,2276. .2302))  
/gene="v1r1"  
/note="v1r1, rpf3, putative transmembrane protein, similar  
to v1r33: MW:35552 (301 aa) fasta scores: opt: 437, E():  
0.00013, 26.5% identity in 302 aa overlap"  
/codon\_start=1  
/protein\_id="CAB96690.1"  
/db\_xref="GI:8953642"  
/translation="MADETYNDSEKRVYVISPKNKEYESVTPSSDRTIVSACNOIRSK  
ISAFSEVTCVKITQYLVHINHEXANODVNDKCEFLYMINVINVPQVDTSTF  
FNDISVYKEOKSSKINICLNMKYINNVLKDLKILMDLFGNFKKEINKEETDC  
SLGECVSYIGSLDKCDKNNTKRCNILEVKNNEAPRIGYCNVORLPVKG  
VLNRSYAVIAFTPIILSSIVFLIKETPILOSWIRPQIVEKKILKLNLOETFEKLOE  
NKNESGTNRNGENLAYNAVHN"  
join(7387. .7404,7544. .8248,8372. .8527)  
/gene="v1r2"  
join(7387. .7404,7544. .8248,8372. .8527)  
/gene="v1r2"  
/note="v1r2, putative transmembrane protein, similar to  
v1r15: MW:36805 (312 aa) fasta scores: opt: 380, E():  
0.00049, 28.0% identity in 296 aa overlap, and to v1r9:  
MW:34499 (294 aa), fasta scores: opt: 377, E(): 0.00053,  
27.2% identity in 290 aa overlap"  
/codon\_start=1  
/protein\_id="CAB96691.1"  
/db\_xref="GI:8953643"  
/translation="MGSESHYHFSNRSYEHESLPLNTNKNVQGHDICEKEMASYI  
YKSTFNLCNKSLNLEYLESEHNDIRSEOGTLVLMYKRELKNVKNKSHLDY  
NNLLEICEFYISYNTGITYOSVNRADNFEILKLDLYKKPKIKDYDCENTYKCA  
KNCPDYKYEIEDCKKKNRDPNCNLENFNRPNFYISSEPECKRKDLIPIRNTYTK  
SVPLISLVTVLSVTFEFLIKLTPFGAMFNRDKNGKNKNSNIAENPEPLHASRI  
SRMSDIKTKYKISYSK"  
join(11806. .11817,11933. .12763,12882. .13034)  
/gene="v1r3"  
join(11806. .11817,11933. .12763,12882. .13034)  
/gene="v1r3"  
/note="v1r3, rpf1, putative transmembrane protein, member  
of rpf subfamily, similar to rpf2, pseudogene 5', v1r31:  
MW:26193 (221 aa) fasta scores: opt: 862, E(): 8.2e-16,  
55.5% identity in 218 aa overlap and rpf2, pseudogene 3',  
and to rpf3, v1r17 MW:46468 (406 aa) fasta scores: opt:  
415, E(): 1.9e-06, 27.7% identity in 393 aa overlap"  
/codon\_start=1  
/protein\_id="CAB96692.1"  
/db\_xref="GI:8953644"  
/translation="MTSISNFEDECKLSAFIKSLDSSPOMQAVIDISYVINSKREOKD  
TVREIYSELESYSIASIOEDNIKSECFYLNMYLNKORKEKILINVDAMVITKLM  
GTLKRPYSCKRQHYEBSVDKRCVDLAVYCNVDELOKHEKPDQAEKFTSYDNF  
NKTIRHYTYFTINAKLDTRNDIDHFMKRSDDTCLHDMATKFEKFTSSQTIYDGG  
TREKIRKESHEHSDRINCVMIDGVPLLEISATINVIPLKYGIIAGSSPFGTISLG  
LYLKRTKHPISLVTRNSSRENKNTKRLKLSHEKEKSKNSKNKDYKFSYNDION"  
join(16650. .16682,16811. .17680,17800. .17964)  
/gene="v1r4"  
join(16650. .16682,16811. .17680,17800. .17964)  
/gene="v1r4"  
/note="v1r4, rpf2, putative transmembrane protein, member of  
rpf subfamily, similar to rpf3, v1r14: MW:40917 (348 aa)  
fasta scores: opt: 878, E(): 8.4e-16, 42.1% identity in  
349 aa overlap, and to rpf1, v1r10: 0.4: v1r10: 0.4: tab

12087:13361 reverse MW:39223 (343 aa) fasta scores: opt:  
597, E(): 5.7e-10, 44.7% identity in 351 aa overlap and to  
rpf4, v1r10: 0.20: v1r10: 0.20: tab 81364:82388 reverse MW:34896  
(299 aa) fasta scores: opt: 441, E(): 8.7e-07, 35.4%  
identity in 339 aa overlap"  
/codon\_start=1  
/protein\_id="CAB96693.1"  
/db\_xref="GI:8953645"  
/translation="WTEKDIFGKLSVETISSDFNSKFFNYILNNDLKKYPRCT  
SLNPKSKDITNTNLICLAOLVKYIKTYTTLNKGDLVDDCLILNVIYSRLVNLIG  
TEDKTVIAPIRGKILVWNSIYANPLNYSYFNKPKDILIPONDNRKRELYDCNV  
YDTIKNTLPIYDOMCEPKYMNVEHSITSLFEKFTLCSKSDQCEFEYNECDOYBNV  
LRTFECNETPMKKAKKEETSEKDKLOJGSGOGEANGMFTNPEFSGGSHLPDGT  
HPAOTGNIILIGVATSTSGALVRFPLGMLNRNGFRRNNHNNRMHMGEGYGLDYA  
PESTNPTPGGGEHEHIGYHPA"  
join(20974. .21000,21140. .21889,22031. .22477,22627. .22788)  
/gene="v1r5"  
join(20974. .21000,21140. .21889,22031. .22477,22627. .22788)  
/gene="v1r5"  
/note="v1r5, hypothetical protein, splicing uncertain,  
low similarity to other v1r10 CDS"  
/codon\_start=1  
/protein\_id="CAB96694.1"  
/db\_xref="GI:8953646"  
/translation="MEEVEERKDYDIFKLDVDEYIIHQVNRYSDEPOANCDISITF  
TNEKNQVSYLCRKVYSEFKIKHNRYSNPDOKKVEYLINPIRHOLOISISKY  
DSVYIHLKNNTYPODKERELEENKIYINERKDSNMLIDLYKNTYGSLEHETEC  
NYFOSFEKENDKLCYRCYFGHDSKLCVNMIFRELVDOKFPNIONCHITYSTSLP  
ELSDVSEKSTJDSDEGFLVKTSTSPGFLQIAGVYKELNELIRQYNDIYAPY  
YDEMKRPMKMLILOFIFKCYRENKNSLSFMEFEKYEKYNENYNDIFADCKI  
GCKSKIYGLKYSCKENYGDLLILEKVNYSYLEOKKYISLSLEWITRAOSLFO  
DGMASRSPYVISMPLPSISIRKEKKRRIPLFEFERSINDYKNNSNMKMKSK  
RKRIRAYQPN"  
join(24943. .24969,25109. .26278,26390. .26551)  
/gene="v1r6"  
join(24943. .24969,25109. .26278,26390. .26551)  
/gene="v1r6"  
/note="v1r6, putative transmembrane protein, similar to  
v1r28: (458 aa), fasta scores: opt: 344, E(): 0.00048,  
26.3% identity in 468 aa overlap"  
/codon\_start=1  
/protein\_id="CAB96695.1"  
/db\_xref="GI:8953647"  
/translation="MAGDHOPGYISYNDVYNAKHEFRYNNRPNDREREKILINKI  
NNSHKMNLKKTFTTLNVLANDLAELFGKHHRCRTINWLNKEVNNINHPDKSY  
EPFIOQFSDERSIITKNNNDORCNNTYINLHKITINDILYGLYDEDKIRSHREDS  
NYGSDPFLSMAASHNVAIDRYEWNTLYOFEEELKILIDMLKESSSSCKISYLAHK  
POYVLRREEEARREAAKORALEEALREERAEKROHDELAERELDROANNTV  
GPTASGEGTPOHTDRSDLRAPITISELDONNORRGYPRSSRGLILDNGHTREGOF  
ITYNEDTRQGEYIEQDGTIKTSNGMGSSGSPRYITTEVFGSDYDPVYVGSAGMG  
ALFLFRTPLPSTFFRGGRAHRIPRSFNQLCAFDPIDNEYNGYIGYPMIDPYG  
AE"  
join(29296. .29313,29447. .30361,30480. .30638)  
/gene="v1r7"  
join(29296. .29313,29447. .30361,30480. .30638)  
/gene="v1r7"  
/note="v1r7, rpf1, putative transmembrane protein, member  
of rpf subfamily, similar to rpf2, v1r29: MW:42155 (362  
aa) fasta scores: opt: 805, E(): 1.4e-07, 39.8% identity  
in 374 aa overlap, and to rpf3, v1r27: MW:42372 (362 aa)  
fasta scores: opt: 487, E(): 0.0004, 26.9% identity in 376  
aa overlap"  
/codon\_start=1  
/protein\_id="CAB96696.1"  
/db\_xref="GI:8953648"  
/translation="MAEILGDEKLEILPTKHNYYLYDNGYFCEKEDIFYNHAERLSS  
DSGOKVEKILKAMQYMYTKRKRKIFENSDMKCFEYVLAIDLINLNDHFTSVILI  
NLYRLINAGAGKICDPIINSYIDNDNPEISDILIDYSDYSYKDLAIPRSCNENY  
KXIQTIVYNNKRELAGICEVEXKRSDDTYCKEKKHPSKNEHNLSTWGNLEGTQLN  
LIGDNGVVEAEQIGKSTGKGAELKTEKVBGRADAEPPVLASLEFGHSMSS  
SSEVEDSSISITSTKILYANVAVAGIYPSLVYIPAGTWINRLEGRTRKMNHNTL  
ADMEILNFSQTSFNFERSRFDISYRV"  
join(34090. .34134,34277. .35083,35199. .35366)





```

* 7279      8126: contig of 848 bp in length
* 8127      8146: gap of unknown length
* 8147      9391: contig of 1245 bp in length
* 9392      9411: gap of unknown length
* 9412      10106: contig of 695 bp in length
* 10107     10126: gap of unknown length
* 10127     11256: contig of 1130 bp in length
* 11257     11276: gap of unknown length
* 11276     12281: contig of 1005 bp in length
* 12281     12301: gap of unknown length
* 12301     13305: contig of 1004 bp in length
* 13305     13325: gap of unknown length
* 13325     14042: contig of 717 bp in length
* 14042     14062: gap of unknown length
* 14062     14871: contig of 809 bp in length
* 14871     14891: gap of unknown length
* 14891     15691: contig of 800 bp in length
* 15691     15711: gap of unknown length
* 15711     17102: contig of 1391 bp in length
* 17102     17122: gap of unknown length
* 17122     18533: contig of 1411 bp in length
* 18533     18553: gap of unknown length
* 18553     19689: contig of 1136 bp in length
* 19689     19709: gap of unknown length
* 19709     21533: contig of 1824 bp in length
* 21533     21553: gap of unknown length
* 21553     23430: contig of 1877 bp in length
* 23430     23450: gap of unknown length
* 23450     25116: contig of 1666 bp in length
* 25116     25136: gap of unknown length
* 25136     26550: contig of 1414 bp in length
* 26550     26570: gap of unknown length
* 26570     28056: contig of 1486 bp in length
* 28056     28076: gap of unknown length
* 28076     28243: contig of 167 bp in length
* 28243     28263: gap of unknown length
* 28263     30115: contig of 1852 bp in length
* 30115     30135: gap of unknown length
* 30135     31923: contig of 1788 bp in length
* 31923     31943: gap of unknown length
* 31943     33125: contig of 1182 bp in length
* 33125     33145: gap of unknown length
* 33145     33955: contig of 810 bp in length
* 33955     33975: gap of unknown length
* 33975     35756: contig of 1781 bp in length
* 35756     35776: gap of unknown length
* 35776     37277: contig of 1501 bp in length
* 37277     37297: gap of unknown length
* 37297     38619: contig of 1322 bp in length
* 38619     38639: gap of unknown length
* 38639     39780: contig of 1141 bp in length
* 39780     39800: gap of unknown length
* 39800     41047: contig of 1247 bp in length
* 41047     41067: gap of unknown length
* 41067     41938: contig of 871 bp in length
* 41938     41958: gap of unknown length
* 41958     43179: contig of 1221 bp in length
* 43179     43199: gap of unknown length
* 43199     44738: contig of 1539 bp in length
* 44738     44758: gap of unknown length
* 44758     46101: contig of 1343 bp in length
* 46101     46121: gap of unknown length
* 46121     47436: contig of 1315 bp in length
* 47436     47456: gap of unknown length
* 47456     49356: contig of 1900 bp in length
* 49356     49376: gap of unknown length
* 49376     51129: contig of 1753 bp in length
* 51129     51149: gap of unknown length
* 51149     53630: contig of 2481 bp in length
* 53630     53650: gap of unknown length
* 53650     56012: contig of 2362 bp in length
* 56012     56032: gap of unknown length
* 56032     57861: contig of 1829 bp in length.

```

```

FEATURES
  source
    1..57861
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="RP8-24A2"
BASE COUNT      15582 a 12043 c 11927 g 15854 t 2455 others
ORIGIN
Query Match      17.7%; Score 34; DB 2; Length 57861;
Best Local Similarity 61.1%; Pred. No. 5.9;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
Oy 3 tggcgccggcgccctgaatgtgcttgacaacatacttcgcatcatttc 62
Db 44537 tggcgctactgttaaccgcccgcgacagatracattttttatgtctttttt 44596
Oy 63 aataccctcgggaaataatgtgaaac 92
Db 44597 gatgacttgggaaatagcttttacttcc 44626

RESULT 14
AC008554 166872 bp DNA PRI 01-JUL-2000
LOCUS Homo sapiens chromosome 19 clone CTC-513N18, complete sequence.
DEFINITION
AC008554
AC008554.7 GI:8886966
KEYWORDS
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 166872)
  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE
    Direct Submission
  JOURNAL
    Unpublished
  AUTHORS
    2 (bases 1 to 166872)
  TITLE
    DOE Joint Genome Institute.
  JOURNAL
    Direct Submission
  AUTHORS
    Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
    Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  REFERENCE
    3 (bases 1 to 166872)
    DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE
    Direct Submission
  JOURNAL
    Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
    Drive, Walnut Creek, CA 94598, USA
  COMMENT
    On Jul 1, 2000 this sequence version replaced g1:7711296.
    Draft Sequence Produced by DOE Joint Genome Institute
    www.jgi.doe.gov
    Finishing Completed at Stanford Human Genome Center
    www.shgc.stanford.edu
    Quality: Phrap Quality >=40 99.7% of Sequence;
    Estimated Total Number of Errors is 2.1.
FEATURES
  source
    1..166872
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="19"
    /clone="CTC-513N18"
BASE COUNT      50087 a 32869 c 34967 g 48949 t
ORIGIN
Query Match      17.7%; Score 34; DB 9; Length 166872;
Best Local Similarity 54.9%; Pred. No. 6.9;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
Oy 43 tctcgattcgctcttttatacttcgggaaatagatgtgaaacccctataaac 102
Db 144050 tctatataattgactctatatttaattcttcagaaatgaaacacattatttaca 143991
Oy 103 gcgggttttcgagaacatgcgctagatcatgtgatgacaacatgactaagcaaat 162

```

Db 14390 GTTTTGGAAATTAAAAATGCTTAAGAAACATTATAAAACTAAAGTAAAAATAAT 14391

QY 163 gc 164

Db 143930 TC 143929

RESULT 15

AC010636  
FOCUS

LOCUS  
DEFINITY

DEFINITION  
ACCESSIO

ACCESSION  
VERSION

## KEYWORDS

**SOURCE**

## ORGANI

**DEPENDENT**

## REFERENCE

AUTHOR	TITLE
...	...

LIFE  
JOURNA

## REFERENCE

**AUTHOR**

**TITLE**

JOURNA

## REFERENCES

REFERENCE  
AUTHOR.

**AUTHOR**  
**TITLE**

JOURNA

**COMMENT**

Drat Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.7% of Sequence;  
Estimated Total Number of Errors is 1.2.

```
source 1..179393
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1p"
        /clone="CTD-2332E11"
BASE COUNT 52588 a 36507 c 35542 g 54756 t
ORIGIN
```

Query Match	17.7%;	Score 34;	DB 9;	Length 179393;
Best Local Similarity	54.9%;	Pred. No. 7;		
Matches 67;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;

43. *tgctgatcgctatttcaataaccttcggggaatagatgtgaaaacctataaac* 102

Db 128506 TGTATAATTGACTCTATTTAATTCTCAGAAAAATTAAC TGACACATTATTACAA 128565

103 gcgggtttcgcagaacatgcgctagtatcatctgatgacaacatggactaagcaaaagt 162

Db 128566 GTTTTTCGAAATTAATAATGCTTAAGAACATTTATAAAACTAAAGTAAAAATTAATT 128625

QY 163 gc 164

Db 128626 TC 128627

Search completed: March 28, 2002, 15:16:11  
Job time: 7935 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 13:16:56 ; Search time 240.01 Seconds

(without alignments)  
685.832 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_1\_192

Perfect score: 192  
Sequence: 1 gctggcgcgcgcgccatga.....tgaccacaagaagatgcttt 192

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :

N.Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	33.4	17.4	50000	21	AAA64140
2	32.6	17.0	278	21	AAC27526
3	30.4	15.8	1569	22	AAH52605
4	30.4	15.8	3204	22	AAH54121
5	29.2	15.2	3077	21	AAA46670
6	29.2	15.2	3092	14	AAO48934
7	29.2	15.2	3126	16	AAO90112
8	29.2	15.2	3202	21	AAC78077
9	29.2	15.2	3253	14	AAO48935
10	29.2	15.2	3324	22	AAH12765
11	29.2	15.2	3405	22	AAH89937

12	29.2	15.2	3448	22	AAH72618
13	29.2	15.2	160552	22	AAH02697
C 14	29	15.1	860	21	AAC74426
C 15	29	15.1	1444	22	AAH34227
C 16	29	15.1	1444	22	AAH05171
C 17	29	15.1	92407	22	AAH28549
C 18	28.6	14.9	16386	22	AAH63695
C 19	28	14.6	2882	22	AAH18508
C 20	27.8	14.5	730	22	AAH21896
C 21	27.8	14.5	1179	22	AAH52400
C 22	27.8	14.5	1257	15	AAO78141
C 23	27.8	14.5	1297	21	AAC48453
C 24	27.8	14.5	1311	21	AAC43114
C 25	27.8	14.5	1603	21	AAC33162
C 26	27.8	14.5	4434	22	AAH54373
C 27	27.8	14.5	5391	19	AAH71729
C 28	27.8	14.5	5407	19	AAV71738
C 29	27.8	14.5	24183	22	AAH21771
C 30	27.6	14.4	285	21	AAH78563
C 31	27.6	14.4	967	21	AAC45504
C 32	27.6	14.4	1623	21	AAH78542
C 33	27.4	14.3	484	22	AAH84040
C 34	27.4	14.3	936	22	AAH58252
C 35	27.4	14.3	936	22	AAH58254
C 36	27.4	14.3	936	22	AAH58257
C 37	27.4	14.3	936	22	AAH58259
C 38	27.4	14.3	936	22	AAH58262
C 39	27.4	14.3	936	22	AAH58255
C 40	27.4	14.3	1143	22	AAH67361
C 41	27.4	14.3	1266	22	AAH67831
C 42	27.4	14.3	1386	22	AAH07063
C 43	27.4	14.3	349980	22	AAH68531
C 44	27.4	14.3	580073	18	AAH58840
C 45	27.2	14.2	750	21	AAC47825

#### ALIGNMENTS

RESULT 1	AAA64140/c	
ID	AAA64140	standard, DNA; 50000 BP.
XX	AAA64140;	
AC	20-DEC-2000	(first entry)
DT		
DE	Nucleotide sequence of a beta-tubulin antigen.	
XX		
KW	Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;	
XX	chronic ear disease; autoimmune disease; ss.	
OS	Homo sapiens.	
XX		
PN	WO200050593-A1.	
XX		
PD	31-AUG-2000.	
XX		
PF	25-FEB-2000; 2000MO-US04795.	
XX		
PR	25-FEB-1999; 99US-0121549.	
XX		
PA	(UYTE-) UNIV TENNESSEE RES CORP.	
XX		
PI	YOO TJ;	
XX		
DR	WPI; 2000-558400/51.	
XX		
PT	New beta-tubulin antigen in the membranous structure of the inner ear,	
PT	reactive with antibodies of patients with Meniere's disease, for	
PT	diagnosing Meniere's disease and distinguishing this disease from other	
XX	autoimmune ear diseases	

PS Claim 3; Page 74-97; 115pp; English.

XX The present sequence encodes a beta-tubulin antigen. The protein is

CC an antigen of the membranous structure of the inner ear protein, and

CC is reactive with antibodies from patients having Meniere's disease.

CC Meniere's disease is a chronic ear disease with unknown etiology.

CC Serum from patients suffering from this disease contain autoantibodies

CC against a 30 kDa cochlear protein antigen. The disease is believed to be

CC an autoimmune disease. The beta-tubulin antigen is useful as a target

CC substance in diagnosing or detecting Meniere's disease and in

CC distinguishing this disease from other autoimmune ear diseases.

XX

SO Sequence 50000 BP; 17281 A; 9480 C; 8791 G; 14448 T; 0 other;

Query Match 17.4%; Score 33.4; DB 21; Length 50000;

Best Local Similarity 52.5%; Pred. No. 0.41;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 42 ttgtctgattcgtctatttcaataccttcggggaatagatgtgaaaccttataaa 101

DB 26376 TTTTCTACCTTTCCTCTTACACGCTACCAACCAATGATCTGTCACGACGAATGAAG 26317

OY 102 cgcgaggttttcgagaagaacatgcctagatcatgtgacacatgagactagaacaaag 161

DB 26316 CCATATTTTACGATTAATATTTCTCTGCCATGATTAATAATGAAATCAGCTAAGCATGAT 26257

OY 162 tgcctgtcccttgaccacaa 180

DB 26256 TTCTTACTGCTGTGACCAA 26238

RESULT 2

AAC27526

ID AAC27526 standard; cDNA; 278 BP.

XX AAC27526;

AC

XX 06-OCT-2000 (first entry)

DT

DE Human secreted protein 5' EST, SEQ ID NO: 31601.

XX

XX Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;

KM gene therapy; chromosome mapping; ss.

KW

XX Homo sapiens.

OS

XX EP1033401-A2.

PN

XX 06-SEP-2000.

PD

XX 21-FEB-2000; 2000EP-0200610.

PF

XX 26-FEB-1999; 99US-0122487.

PR

XX (GEST ) GENSET.

PA

XX Dumas Mline Edwards J, Duclert A, Giordano J;

PI

XX WPI; 2000-500381/45.

DR

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX

XX Claim 1; SEQ ID 31601; 71pp + CD-ROM; English.

PS

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

1

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors.

XX

SO Sequence 278 BP; 82 A; 44 C; 43 G; 108 T; 1 other;

Query Match 17.0%; Score 32.6; DB 21; Length 278;

Best Local Similarity 51.0%; Pred. No. 0.08;

Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 18 tgaatggtcttacaagaacatactgtctgtatctgtctatttcaataccttcgggaa 77

DB 99 tcattgtgattgttttctgtctttagtataaagctgtattcaataattkggggac 158

OY 78 atagatgtgaaacccctataaagcggttttcgagaacatgcctagtatcatgt 137

DB 159 atatatgtcatctcacatatatacagatgtgtatcaggttgacattccaaagtgcactg 218

OY 138 atgacaacatgactaagaacaaagtctgt 168

DB 219 ctgaglttagaggtatcatgtattgttaattc 249

RESULT 3

AAH52605

ID AAH52605 standard; DNA; 1569 BP.

XX AAH52605;

AC

XX 03-SEP-2001 (first entry)

DT

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:603.

XX

XX Staphylococcus epidermidis SRI strain; infection; diagnosis;

KM vaccination; endocarditis; ds.

KW

XX Staphylococcus epidermidis.

OS

XX WO200134809-A2.

PN

XX 17-MAY-2001.

PD

XX 09-NOV-2000; 2000WO-US30782.

PF

XX 09-NOV-1999; 99US-0164258.

PR

XX (GLAXO ) GLAXO GROUP LTD.

PA

XX Kimmery MJ;

PI

XX WPI; 2001-316495/33.

DR

XX P-PSDB; AAG81755.

DR

XX Nucleic acids encoding polypeptides from staphylococcus epidermidis,

PT useful for vaccinating against infections, e.g. endocarditis -

PT

XX

XX Claim 8; Page 195; 2188pp; English.

PS

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their



PS Disclosure; Fig 7C; 170pp; English.

XX AAA46668-79 represent differentially expressed human genes, associated  
 CC with disease states and disorders. The specification describes  
 CC methods preventing, diagnosing and treating cardiac, kidney and  
 CC inflammatory diseases associated with inappropriate expression of  
 CC differentially expressed cardiac, kidney and inflammatory genes  
 CC (e.g. AAA46668-79). These genes include I-80, prostacyclin-stimulating  
 CC factor, 1sf-2, tissue specific mRNA, insulin-like growth factor  
 CC binding protein 6, OSF-1, gas-1, YMP, ERG2, pre-B cell stimulating  
 CC factor homologue (SOF1a), peripheral benzodiazepine receptor, and  
 CC cellular ligand of annexin II (p11), respectively. These diseases  
 CC include congenital heart failure, dilated congestive cardiomyopathy,  
 CC hypertrophic cardiomyopathy, restrictive cardiomyopathy, mitral valve  
 CC disease, aortic valve disease, tricuspid valve disease, angina pectoris,  
 CC myocardial infarction, cardiac arrhythmia, pulmonary hypertension,  
 CC arterial hypertension, renovascular hypertension, arteriosclerosis,  
 CC atherosclerosis and/or cardiac tumours.

XX Sequence 3077 BP; 1032 A; 571 C; 632 G; 842 T; 0 other;

QY Query Match 15.2%; Score 29.2; DB 21; Length 3077;  
 Best Local Similarity 59.8%; Pred. No. 3.7;  
 Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccttataaagcggttttcgcagaacatgcgctagatcatgatgacaacatg 148  
 ||||| ||||| | ||||| || | | ||||| ||||| ||||| ||||| |||||  
 Db 616 aacctatcctaattggtgtgtcactgttaattgtgtcgaatcatcatggaacag 675  
 | |||| | |||||

OY 149 gactaagcaaaagtgtgtcc 170  
 | |||| | |||||

Db 676 attgcaacaatggtgtgtcc 697

RESULT 6  
 AAQ48934  
 ID AAQ48934 standard; cDNA to mRNA; 3092 BP.

XX AAQ48934;  
 XX 15-APR-1994 (first entry)  
 XX hosf-2pl.  
 XX Bone-related protein; bone; diagnosis; disease; growth factor;  
 XX cell adhesion; guiding; induction; metabolic bone disease; ss.  
 XX Homo sapiens (placenta).

XX Key Location/Qualifiers  
 XX CDS 38..2377  
 XX /tag- a  
 XX /product- OSF-2  
 XX sig-peptide 38..101  
 XX mat-peptide 102..2374  
 XX /\*tag- b  
 XX /\*tag- c

XX EP562508-A.  
 XX 29-SEP-1993.  
 XX 22-MAR-1993; 93EP-0104650.  
 XX 27-MAR-1992; 92JP-0071501.  
 XX (FARH ) HOECHST JAPAN LTD.  
 XX Amano E, Kikuno R, Otawara-Hamamoto Y, Takeshita S,  
 XX Tezuka K;  
 XX WPI; 1993-304910/39.

DR P-PSDB; AAR41868.

XX Mammalian OSF-2 protein - obtd. from bone tissue, useful for  
 PT diagnosis and treatment of metabolic bone disease  
 XX Claim 6; Page 14-18; 58pp; English.

XX OSF-2 plays an important role in the formation of bone, by acting as  
 CC a growth factor or adhesion or "guiding" protein to attract cells  
 CC to the site of bone induction. In a pharmaceutical compsn. It can  
 CC be applied in metabolic bone diseases.

XX Sequence 3092 BP; 1035 A; 573 C; 638 G; 846 T; 0 other;

QY Query Match 15.2%; Score 29.2; DB 14; Length 3092;  
 Best Local Similarity 59.8%; Pred. No. 3.7;  
 Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccttataaagcggttttcgcagaacatgcgctagatcatgatgacaacatg 148  
 ||||| ||||| | ||||| || | | ||||| ||||| ||||| ||||| |||||  
 Db 626 aacctatcctaattggtgtgtcactgttaattgtgtcgaatcatcatggaacag 685  
 | |||| | |||||

OY 149 gactaagcaaaagtgtgtcc 170  
 | |||| | |||||

Db 686 attgcaacaatggtgtgtcc 707

RESULT 7  
 AAQ90112  
 ID AAQ90112 standard; cDNA; 3126 BP.

XX AAQ90112;  
 XX 05-NOV-1995 (first entry)  
 XX TCI gene.  
 XX Tumour marker; invasive; metastatic; cancer; ss.  
 XX Homo sapiens.

XX Key Location/Qualifiers  
 XX CDS 43..2376  
 XX /\*tag- a

XX W09511923-A.  
 XX 04-MAY-1995.  
 XX 31-OCT-1994; 94WO-US12502.  
 XX 29-OCT-1993; 93US-0146488.  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX Bao S, Chen LB, Liu Y;  
 XX WPI; 1995-178826/23.  
 XX P-PSDB; AAR74302.

XX New tumour marker TCI, corresp. DNA and monoclonal antibody - for  
 PT detecting, preventing and treating tumours, esp. in breast, colon  
 PT and gastrointestinal tract cancer.

XX Disclosure; Fig 4; 84pp; English.

XX The sequence is that of the TCI gene which encodes the TCI tumour  
 CC marker protein. The gene and its product may be used to detect  
 CC tumours in blood, urine or sputum. Inhibitors of TCI are used to  
 CC treat late stage cancers and for preventing tumour cell metastasis.  
 CC See also AAQ90113-25.





CC chemical compounds as potential drugs.



Query Match	15.2%	Score 29.2	DB 22	Length 160552
3'UTR	FT	(GST-4alpha)	49129..49746	
FT	FT	/tag=1	83257..83347	
FT	FT	/tag=m	83348..96412	
FT	FT	/tag=n	96413..96484	
FT	FT	/tag=o	96485..98456	
FT	FT	/tag=p	98457..99968	
FT	FT	/tag=q	99969..100000	
FT	FT	/tag=r	100001..100000	
FT	FT	/tag=s	100001..100000	
FT	FT	/tag=t	100001..100000	
FT	FT	/tag=u	100001..100000	
FT	FT	/tag=v	100001..100000	
FT	FT	/tag=w	100001..100000	
FT	FT	/tag=x	100001..100000	
FT	FT	/tag=y	100001..100000	
FT	FT	/tag=z	100001..100000	
FT	FT	/tag=aa	100001..100000	
FT	FT	/tag=ab	100001..100000	
FT	FT	/tag=ac	100001..100000	
FT	FT	/tag=ad	100001..100000	
FT	FT	/tag=ae	100001..100000	
FT	FT	/tag=af	100001..100000	
FT	FT	/tag=ag	100001..100000	
FT	FT	/tag=ah	100001..100000	
FT	FT	/tag=ai	100001..100000	
FT	FT	/tag=aj	100001..100000	
FT	FT	/tag=ak	100001..100000	
FT	FT	/tag=al	100001..100000	
FT	FT	/tag=am	100001..100000	
FT	FT	/tag=an	100001..100000	
FT	FT	/tag=ao	100001..100000	
FT	FT	/tag=ap	100001..100000	
FT	FT	/tag=aq	100001..100000	
FT	FT	/tag=ar	100001..100000	
FT	FT	/tag=as	100001..100000	
FT	FT	/tag=at	100001..100000	
FT	FT	/tag=au	100001..100000	
FT	FT	/tag=av	100001..100000	
FT	FT	/tag=aw	100001..100000	
FT	FT	/tag=ax	100001..100000	
FT	FT	/tag=ay	100001..100000	
FT	FT	/tag=az	100001..100000	
FT	FT	/tag=ba	100001..100000	
FT	FT	/tag=bb	100001..100000	
FT	FT	/tag=bc	100001..100000	
FT	FT	/tag=bd	100001..100000	
FT	FT	/tag=be	100001..100000	
FT	FT	/tag=bf	100001..100000	
FT	FT	/tag=bg	100001..100000	
FT	FT	/tag=bh	100001..100000	
FT	FT	/tag=bi	100001..100000	
FT	FT	/tag=bj	100001..100000	
FT	FT	/tag=bk	100001..100000	
FT	FT	/tag=bl	100001..100000	
FT	FT	/tag=bm	100001..100000	
FT	FT	/tag=bn	100001..100000	
FT	FT	/tag=bo	100001..100000	
FT	FT	/tag=bp	100001..100000	
FT	FT	/tag=bq	100001..100000	
FT	FT	/tag=br	100001..100000	
FT	FT	/tag=bs	100001..100000	
FT	FT	/tag=bt	100001..100000	
FT	FT	/tag=bu	100001..100000	
FT	FT	/tag=bv	100001..100000	
FT	FT	/tag=bw	100001..100000	
FT	FT	/tag=bx	100001..100000	
FT	FT	/tag=by	100001..100000	
FT	FT	/tag=bz	100001..100000	
FT	FT	/tag=ca	100001..100000	
FT	FT	/tag=cb	100001..100000	
FT	FT	/tag=cc	100001..100000	
FT	FT	/tag=cd	100001..100000	
FT	FT	/tag=ce	100001..100000	
FT	FT	/tag=cf	100001..100000	
FT	FT	/tag=cg	100001..100000	

Best Local Similarity 52.5%; Pred. No. 21;  
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Oy 2 ctggagcgccgcgccaagaatgttgcaacaagaataatctgtcgttatttt 61  
Db 105426 ctggcctcgggaagctcgaagggaacaaacactgatggcctgtatggtcataga 105485

Oy 62 caataccctcgggggaatatagtgcgaacacctataaacgcggatttcgcaanaa 121  
Db 105486 aacagctgttcctcctgaatagtctcagagaccacactatcaagaagcctgttgggaatcaa 105545

Oy 122 tg 123  
Db 105546 tg 105547

RESULT 14  
AAC74426/c  
ID AAC74426 standard; cDNA; 860 BP.

XX AAC74426:  
AC AACT74426:  
XX 05-FEB-2001 (first entry)  
DT XX  
DE Human secreted protein gene .31 SEQ ID NO:41.  
XX

KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
KW antirheumatic; antiproliferative; cytosolic; cardiant; vasotropic;  
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
KW fungicide; pathomimological; vulnary; gene therapy; angiogenesis;  
KW autoimmune disease; hyperproliferative disorder; infection; skin aging;  
KW cardiovascular disorder; cerebrovascular disorder; ocular disorder;  
KW nervous system disorder; wound healing; food additive; preservative; ss.  
XX  
OS Homo sapiens.  
PN MN  
NM WO200058496-A1.  
PD 05-OCT-2000.  
PF 23-MAR-2000; 2000WO-USO7722.  
PR 26-MAR-1999; 99US-0126508.  
PR 07-JAN-2000; 2000US-0174871.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Ruben SM, Komatsoulis G;  
PI WPI: 2000-594645/56.  
DR P-PSDB; AAB40181.

XX New nucleic acid molecules encoding 50 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX Claim 1; Page 324; 408pp; English.

XX The polynucleotide sequences given in AAC74396 to AAC74445 encode the  
CC human secreted proteins given in AAB40151 to AAB40200. AAB40201 to  
CC AAB40236 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;  
CC cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
CC and vulnary. The polynucleotides and polypeptides can be used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. Disorders which are diagnosed or treated include  
CC autoimmune diseases, hyperproliferative disorders, cardiovascular

CC disorders, cerebrovascular disorders, angiogenesis, nervous system  
CC disorders, infections caused by bacteria, viruses and fungi and ocular  
CC disorders. The polypeptides can also be used to aid wound-healing and,  
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
CC maintain organs before transplantation, for supporting cell culture of  
CC primary tissues, to regenerate tissues and in chemotaxis. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities, fat content, lipid, protein,  
CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
CC components. AAC74387 to AAC74395 and AAB40150 represent sequences used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 860 BP; 293 A; 161 C; 128 G; 278 T; 0 other;

Query Match 15.1%; Score 29; DB 21; Length 860;  
Best Local Similarity 57.0%; Pred. No. 2.5;  
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;  
OY 76 aaatagatgtaaaacccctataaaacgaggttgcgaagaacatgcgtagtatcat 135  
DB 266 AATATATAGGAATCTCTGTTAAGCACCTCCCGTGAAGTAATGAGAGAGATCAT 207  
OY 136 tgaatgacaactgactaagcaaaagtcttgt 168  
DB 206 CATTGACTACATATATATTAAGAAATAAGTCTTTT 174

## RESULT 15

AAH34227/C

ID AAH34227 standard; CDNA; 1444 BP.

XX AAH34227;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding CDNA SEQ ID NO:1309.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; ss.

XX Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX P-PSDB; AAG74822.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides;  
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 3046; 9803pp; English.

CC AAH32943 to AAH37195 and AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAB77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 1444 BP; 487 A; 202 C; 227 G; 527 T; 1 other;

Query Match 15.1%; Score 29; DB 22; Length 1444;  
Best Local Similarity 52.0%; Pred. No. 3.1;  
Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
OY 56 tatttcaatacccttcggggaatagatgtaaaacccctataaaacgaggttgcga 115  
DB 191 TATTACTGAACCTTCCATCTACTACAGAGTCTGTCTTTAAATATTATTAGTTACAA 132  
OY 116 gaacatgctgtagatcatcattgatgacacatggaactgaagaaagtctgtccctga 175  
DB 131 GAATCATTTCTAATGAGTATGATGATTAATGACCAATATATACAGGTGAGTATCCCTTAC 72  
OY 176 cccaa 180  
DB 71 TCAAA 67

Search completed: March 28, 2002, 15:15:31  
Job time: 7115 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 13:05:56 ; Search time 112.49 Seconds  
(without alignments)  
386.557 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_1\_192

Perfect score: 192  
Sequence: 1 gctgggcgcgcgcgcacatga.....tgaccacaagaagatgcttt 192

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents\_NA:\*

1: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/lna/5A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/lna/5B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/lna/PCRTUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/lna/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.2	15.2	3092	1 US-08-426-627-3	Sequence 3, App11
2	29.2	15.2	3126	2 US-08-477-396A-3	Sequence 3, App11
3	29.2	15.2	3253	1 US-08-426-627-5	Sequence 5, App11
4	27.8	14.5	730	4 US-08-943-731-128	Sequence 128, App1
5	27.8	14.5	1257	1 US-08-930-154-1	Sequence 1, App11
6	27.8	14.5	24183	4 US-08-943-731-3	Sequence 3, App11
7	27	14.1	1009	6 5223394-8	Patent No. 5223394
8	27	14.1	1244	2 US-08-204-288-3	Sequence 3, App11
9	27	14.1	3182	1 US-08-188-582-12	Sequence 12, App1
10	27	14.1	3182	1 US-08-646-715-12	Sequence 12, App1
11	27	14.1	3852	1 US-08-306-546C-1	Sequence 1, App11
12	27	14.1	3852	2 US-08-530-524A-1	Sequence 2, App11
13	26.2	13.6	1610	4 US-09-347-803-19	Sequence 19, App1
14	26.2	13.6	2801	3 US-08-747-221B-30	Sequence 30, App1
15	26.2	13.6	2801	3 US-08-747-221B-32	Sequence 32, App1
16	26.2	13.6	2801	4 US-09-005-051-30	Sequence 30, App1
17	26.2	13.6	2801	4 US-09-005-051-32	Sequence 32, App1
18	26.2	13.6	2836	3 US-08-747-221B-24	Sequence 24, App1
19	26.2	13.6	2836	4 US-08-747-221B-26	Sequence 26, App1
20	26.2	13.6	2836	4 US-09-005-051-24	Sequence 24, App1
21	26.2	13.6	2836	4 US-09-005-051-26	Sequence 26, App1
22	25.8	13.4	885	2 US-08-222-719-7	Sequence 7, App11
23	25.8	13.4	885	2 US-08-470-925-7	Sequence 7, App11
24	25.8	13.4	885	2 US-08-471-613-7	Sequence 7, App11
25	25.8	13.4	885	2 US-08-471-613-7	Sequence 7, App11
26	25.6	13.3	735	3 US-09-193-191-3	Sequence 3, App11
27	25.6	13.3	1062	4 US-09-371-913A-5	Sequence 5, App11

28	25.6	13.3	2172	1 US-08-158-232-50	Sequence 50, App1
29	25.6	13.3	2172	1 US-08-611-928-50	Sequence 50, App1
30	25.6	13.3	2172	2 US-09-173-891-50	Sequence 50, App1
31	25.6	13.3	3267	3 US-08-257-963B-12	Sequence 12, App1
32	25.6	13.3	3267	5 PCT-US95-07201-12	Sequence 12, App1
33	25.6	13.3	3997	5 US-08-947-823-2	Sequence 2, App11
34	25.6	13.3	22481	5 PCT-US95-07201-43	Sequence 43, App1
35	25.6	13.3	51952	3 US-08-947-823-1	Sequence 1, App11
36	25.4	13.2	1200	3 US-08-854-531-5	Sequence 5, App11
37	25.4	13.2	1200	5 PCT-US95-13552-5	Sequence 5, App11
38	25.2	13.1	420	1 US-08-680-726A-65	Sequence 65, App1
39	25.2	13.1	420	3 US-09-092-409-65	Sequence 65, App1
40	25.2	13.1	2044	1 US-08-680-726A-63	Sequence 63, App1
41	25.2	13.1	2044	1 US-08-680-726A-64	Sequence 64, App1
42	25.2	13.1	2044	3 US-09-092-409-63	Sequence 63, App1
43	25.2	13.1	2044	3 US-09-092-409-64	Sequence 64, App1
44	25.2	13.1	3758	3 US-08-323-477-1	Sequence 1, App11
45	25.2	13.1	8532	1 US-08-452-655B-1	Sequence 1, App11

## ALIGNMENTS

RESULT 1  
US-08-426-627-3  
; Sequence 3, Application US/08426627  
; Patent No. 5756664  
; GENERAL INFORMATION:  
; APPLICANT: Amano, Egon  
; APPLICANT: Okawara-Hamamoto, Yoko  
; APPLICANT: Kikuno, Reiko  
; APPLICANT: Takeshita, Sunao  
; APPLICANT: Tezuka, Kenichi  
; TITLE OF INVENTION: No. 5756664e1 Protein with Bone Formation  
; TITLE OF INVENTION: Ability and Process for Its Production.  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,627  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/036,841  
; FILING DATE: 25-MAR-1993  
; APPLICATION NUMBER: JP 4-71501  
; FILING DATE: 27-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hammond, Alan W.  
; REGISTRATION NUMBER: 35,178  
; REFERENCE/DOCKET NUMBER: 02481-1285-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3092 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; ORIGINAL SOURCE:

ORGANISM: Homo sapiens  
TISSUE TYPE: Placenta  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(38..2375)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: join(101..2375)  
US-08-426-627-3

Query Match 15.2%; Score 29.2; DB 1; Length 3092;  
Best Local Similarity 59.8%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccctataaagcggttttcgcagaacatgcgctagtcatcattgatacacatg 148  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 626 AACCATATATCCTAATGCGGTTCACGTTAATTGCTCGAATCATCATGGGAACG 685  
QY 149 gactaagcaaaagtgcctgtcc 170  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 686 ATTGCAACAATAATGCTGTGTCC 707

RESULT 2  
US-08-477-396A-3  
Sequence 3, Application US/08477396A  
Patent No. 5872235  
GENERAL INFORMATION:  
APPLICANT: Chen, Ian Bo  
APPLICANT: Bao, Shideng  
APPLICANT: Liu, Yuan  
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF  
NUMBER OF SEQUENCES: 19  
TITLE OF INVENTION: ISOLATING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,396A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/146,488  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: US 08/448,388  
FILING DATE: 28-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/12502  
FILING DATE: 31-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Heine, Holliday C.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: DFCI-333BX  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ. ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3126 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..2376  
US-08-477-396A-3

Query Match 15.2%; Score 29.2; DB 2; Length 3126;  
Best Local Similarity 59.8%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccctataaagcggttttcgcagaacatgcgctagtcatcattgatacacatg 148  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 649 AACCATATATCCTAATGCGGTTCACGTTAATTGCTCGAATCATCATGGGAACG 708  
QY 149 gactaagcaaaagtgcctgtcc 170  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 709 ATTGCAACAATAATGCTGTGTCC 730

RESULT 3  
US-08-426-627-5  
Sequence 5, Application US/08426627  
Patent No. 5756664  
GENERAL INFORMATION:  
APPLICANT: Amann, Egon  
APPLICANT: Otawara-Hamamoto, Yoko  
APPLICANT: Kikuno, Reiko  
APPLICANT: Takeshita, Sunao  
APPLICANT: Tezuka, Kenichi  
TITLE OF INVENTION: No. 5756664el Protein with Bone Formation  
NUMBER OF SEQUENCES: 24  
TITLE OF INVENTION: Ability and Process for Its Production.  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,627  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,841  
FILING DATE: 25-MAR-1993  
APPLICATION NUMBER: JP 4-71501  
FILING DATE: 27-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hammond, Alan W.  
REGISTRATION NUMBER: 35,178  
REFERENCE/DOCKET NUMBER: 02481-1285-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3253 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
CELL TYPE: osteosarcoma

```

:      FEATURE:      CDS
:      NAME/KEY:      join(32...2540)
:      LOCATION:
:      FEATURE:
:      NAME/KEY:      mat_peptide
:      LOCATION:      join(97...2540)
:      :
US-08-426-627-5

```

Query Match	15.28;	Score 29.2;	DB 1;	length 3253;
Best Local Similarity	59.8%;	Pred. No. 1.1;		
Matches 49;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;

Oy	89	aacctcttaaaacgcgaggttttcgcgaacaacatgcgcctgatcatcattggtagaacatg	148
Db	620	AACCATTAATCCTAATGGGCTGTGCAGTGTAATTGTCTCGAATCAATCCATGGGAACCAG	679
Oy	149	gactaagaacaaagtgccttgcacc	170
Db	680	ATTGCACAACAAATGGTGTGTGCC	701

RESULT 4  
 US-08-943-731-128/c  
 Sequence 128, Application US/08943731  
 Patent No. 6265157  
 GENERAL INFORMATION:  
 APPLICANT: PROCKOP, DARWIN J.  
 APPLICANT: SPOTILA, LORETTA D.  
 APPLICANT: DELTAS, CONSTANTINOS D.  
 APPLICANT: SEREDA, LARISSA  
 APPLICANT: LARSON, ANDREA W.  
 APPLICANT: PACK, MICHAEL  
 APPLICANT: COLIGE, ALAIN  
 APPLICANT: EARLY, JAMES  
 APPLICANT: KORRKO, JARMO  
 APPLICANT: ALA-KORRKO, LEENA, et al.  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING  
 TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES  
 NUMBER OF SEQUENCES: 666  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
 STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND  
 STREET: FLR.  
 CITY: PHILADELPHIA  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-7086  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/943,731  
 FILING DATE: 03-OCT-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/212,322  
 FILING DATE: 14-MAR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/803,628  
 FILING DATE: 03-DEC-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: DOYLE LEARY Ph.D., KATHRYN  
 REGISTRATION NUMBER: 36,317  
 REFERENCE/DOCKET NUMBER: 9598-27  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 215-965-1284  
 TELEFAX: 215-567-2991  
 TELEX: 831-494  
 INFORMATION FOR SEQ ID NO: 128:  
 SEQUENCE CHARACTERISTICS:

```

; LENGTH: 730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-128

```

Query Match	14.5%	Score 27.8;	DB 4;	Length 730;
Best Local Similarity	65.1%	Pred. No. 1.8		
Matches 41;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;

Oy	130	tatcatgatgacacacatgagcttaagcaaaagtctgtccctgcacccaagaagatgc	189
Db	374	TTTTGAGGATGTCAGACCTCTGACCAAGCTCAGCTGTATCCCTCGCCCTCTGAAAGCTGC	315
Oy	190	ttt	192
Db	314	TCT	312

```

RESULT 5
US-08-330-154-1
: Sequence 1, Application US/08330154
: Patent No. 5587307
: GENERAL INFORMATION:
: APPLICANT: Alborn Jr., William E
: APPLICANT: Hoskins, Joann
: APPLICANT: Skatrud, Paul L
: APPLICANT: Unal, Serhat
: TITLE OF INVENTION: FEMA GENE OF STAPHYLOCOCCUS EPIDERMIDIS,
: TITLE OF INVENTION: FEMA PROTEIN, AND VECTORS AND MICROORGANISMS COMPRISING
: TITLE OF INVENTION: THE FEMA GENE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Patent Division/AEH
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/330,154
: FILING DATE: 27-OCT-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/208,925
: FILING DATE: 09-MAR-1994
: APPLICATION NUMBER: US/08/057,163
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hamilton, Amy E
: REGISTRATION NUMBER: 33,894
: REFERENCE/DOCKET NUMBER: X-8894
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3169
: TELEFAX: 317-276-1294
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1257 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1257
:
: US-08-330-154-1

```



Query Match	14.5%	Score 27.8	DB 1	Length 1257
Best Local Similarity	62.0%	Pred. No. 2.2		
Matches 44	Conservative 0	Mismatches 27	Indels 0	Gaps 0
QY	89	aaccttaaaagcgggttttcgcagaacatgctagatcatcatgatgaacacatg	148	
Db	654	AAACCTTTAAAGACCGTGTTTAGTACCATGACCTATTATTAACCTTGTATGATATATA	753	
QY	149	gactaagcaaa	159	
Db	754	GAGGAACTTAAA	764	

```

RESULT 6
US-08-943-731-3/c
Sequence 3, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KORRKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHMARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 24183 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-3

```

```

Query Match      14.5%; Score 27.8; DB 4, Length 24183;
Best Local Similarity 65.1%; Pred. No. 7.1;
Matches 41; Conservative 0; Mismatches 22; Indels 0; Gaps 0.

Oy 130 tatcatgatgaacaacatgygactaaagcaaatgcttgcctctgaccacaagaagatgc 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10606 TTTTCGAAGTGCACACCTTTGACCAAGCTCAGCTGGCTATCCCTCCCTCTGAAGGCTCC 10547

Oy 190 ttt 192
    |||
Db 10546 TCT 10544

```

```

RESULT 7
5223394-8
Patent No. 5223394
APPLICANT: WALLNER, BARBARA
TITLE OF INVENTION: RECOMBINANT DNA MOLECULE COMPRISING
LYMPHOCYTE FUNCTION-ASSOCIATED ANTIGEN 3 PHOSPHATIDYLINOSITOL
LINKAGE SIGNAL SEQUENCE
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/335,688
FILING DATE: 10-APR-1989
SEQ ID NO: 8
LENGTH: 1009
5223394-8

```

Query Match	14.18	Score 27	DB 6	Length 1009
Best Local Similarity	57.88	Pred. No. 3.8		
Matches 48	Conservative	0	Mismatches 35	Indels 0
			Gaps	
Oy	74	ggaatagatctggaacccctataaagcggttttcgagaaacatgcgtatgac	133	
Db	777	ggaagtgctcttaaacgggttacagacagacagatgctattaccataccattagc	836	
Oy	134	atgatgacaacatgactgaac	156	
Db	837	agtaattacacacatgattctgc	859	

RESULT 8  
US-08-204-288-3/c  
Sequence 3, Application US/08204288  
Patent No. 5959178  
GENERAL INFORMATION:  
APPLICANT: VAN DOORSSELAERE, Jan  
APPLICANT: FRITIG, Bernard J.M.  
APPLICANT: INZE, Dirk G.  
APPLICANT: JOUANIN, Lise  
APPLICANT: KNIGHT, Mary E.  
APPLICANT: VAN MONTAGU, Marc  
APPLICANT: LEGRAND, Michel  
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN  
TITLE OF INVENTION: PLANTS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
STREET: 1100 New York Avenue, N.W.  
CITY: Washington  
STATE: D. C.  
COUNTRY: U.S.A.  
ZIP: 20005-3518  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/204,288  
FILING DATE: 10-MAR-1994



sequence 19, Application 05/0934/803  
; Patent No. 6274379  
: GENERAL INFORMATION:

APPLICANT: Famodu, Iayo O.  
APPLICANT: Hite, Bill  
APPLICANT: Kinney, Tony  
APPLICANT: Orozco, Buddy  
TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes  
FILE REFERENCE: BR-1176  
CURRENT APPLICATION NUMBER: US/09/347,803  
CURRENT FILING DATE: 1999-07-02  
EARLIER APPLICATION NUMBER: 60/092,952  
EARLIER FILING DATE: July 15, 1998  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 19  
LENGTH: 1610  
TYPE: DNA  
ORGANISM: Glycine max  
US-09-347-803-19

Query Match 13.6%; Score 26.2; DB 4; Length 1610;  
Best Local Similarity 79.5%; Pred. No. 8.8;  
Matches 31; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 72 ggggaatagatgtgaaacccctataaagcggttt 110  
DB 1058 ggcacaattgattgaaacccctataaagcggttt 1096

RESULT 14  
US-08-747-221B-30/C  
Sequence 30, Application US/08747221B  
Patent No. 6063610  
GENERAL INFORMATION:

APPLICANT: Silver, Gary W.  
TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid  
TITLE OF INVENTION: Molecules, Proteins and Uses Thereof  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 2801 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 99..1886

US-08-747-221B-30

Query Match 13.6%; Score 26.2; DB 3; Length 2801;  
Best Local Similarity 52.3%; Pred. No. 11;  
Matches 58; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 60 ttcaatccttcgggggaatagatgtgaaacccctataaagcggttttcgcagaaa 119  
DB 2165 ttaattggcattgcatatgaaatgctgtacccctacaaaatttgccgatttgctaaaa 2106  
QY 120 catgcgtatgatactatgatacacacatgagactaagcaaaagtctgtcc 170  
DB 2105 GAAGAAAAAGCTTCATATATGCGCATGTGCTATATAAAGTGTTC 2055

RESULT 15  
US-08-747-221B-32  
Sequence 32, Application US/08747221B  
Patent No. 6063610  
GENERAL INFORMATION:

APPLICANT: Silver, Gary W.

TITLE OF INVENTION: No. 6063610el Carboxylesterase Nucleic Acid

TITLE OF INVENTION: Molecules, Proteins and Uses Thereof

NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/747,221B

FILING DATE: No. 6063610ember 12, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: FC-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:

LENGTH: 2801 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-747-221B-32

Query Match 13.6%; Score 26.2; DB 3; Length 2801;  
Best Local Similarity 52.3%; Pred. No. 11;  
Matches 58; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 60 ttcaatccttcgggggaatagatgtgaaacccctataaagcggttttcgcagaaa 119  
DB 637 ttaattggcattgcatatgaaatgctgtacccctacaaaatttgccgatttgctaaaa 696  
QY 120 catgcgtatgatactatgatacacacatgagactaagcaaaagtctgtcc 170  
DB 697 GAAGAAAAAGCTTCATATATGCGCATGTGCTATATAAAGTGTTC 747.

Search completed: March 28, 2002, 15:10:40  
Job time: 7484 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 12:36:06 ; Search time 2300.01 Seconds  
(without alignments)  
897.035 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_1\_192

Perfect score: 192  
Sequence: 1 gctggggcgcgcgcgcacatga.....tgaccacaagaagatgcttt 192

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_estnum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	36.4	19.0	749	13	BH039236
C 2	34.6	18.0	852	11	BG542199
C 3	34.2	17.8	600	13	AZ964976
C 4	33.8	17.6	945	13	CNS079HE
C 5	33.6	17.5	845	13	CNS0757U
C 6	33.2	17.3	611	13	AO017410
C 7	33.2	17.2	523	13	AQ268787
C 8	32.8	17.1	957	13	AZ931427
C 9	32.6	17.0	552	13	CNS04NR7
C 10	32.6	17.0	958	13	BG062885
C 11	32.6	17.0	958	13	CNS0480J
C 12	32.6	17.0	1101	13	CNS007P

13	32.6	17.0	2062	12	AK005440
14	32.6	17.0	2366	12	AK004888
15	32.4	16.9	510	13	AO787593
16	32.2	16.8	542	13	AO564194
C 17	32.2	16.8	746	13	AZ900491
C 18	32.2	16.7	439	13	AZ099361
C 19	31.6	16.5	291	10	BB485889
C 20	31.4	16.4	236	10	BB083723
C 21	31.4	16.4	1511	11	BF234077
C 22	31.2	16.2	434	11	BF547099
C 23	31.2	16.2	650	10	AV733876
C 24	31.2	16.2	677	13	AO196619
C 25	31.2	16.2	973	13	CNS02ETM
C 26	31.1	16.1	621	10	AV733687
C 27	31.1	16.1	680	13	AO9227062
C 28	30.8	16.0	769	11	BG068661
C 29	30.6	15.9	318	10	AV296697
C 30	30.6	15.9	320	11	BI202015
C 31	30.6	15.9	354	10	AW443473
C 32	30.6	15.9	613	10	AI779343
C 33	30.6	15.9	615	10	AI779344
C 34	30.6	15.9	1101	13	CNS00BLA
C 35	30.4	15.8	408	10	BES81870
C 36	30.4	15.8	444	13	AO153210
C 37	30.4	15.8	445	10	BES79806
C 38	30.4	15.8	445	11	BG227023
C 39	30.4	15.8	617	13	AO486794
C 40	30.4	15.8	735	13	AO384366
C 41	30.4	15.8	777	11	BG849734
C 42	30.4	15.8	865	11	BF246893
C 43	30.2	15.7	471	11	BG047104
C 44	30.2	15.7	532	11	W81666
C 45	30.2	15.7	565	13	AO534021

#### ALIGNMENTS

RESULT 1  
LOCUS BH039236/c  
DEFINITION RPT-24-254N24.TV RPT-24 Mus musculus genomic clone RPT-24-254N24  
ACCESSION BH039236  
VERSION BH039236.1 GI:14816964  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvatslsdeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.  
TITLE Mouse BAC End Sequences from Library RPT-24  
JOURNAL Unpublished (1999)  
COMMENT Other-GSS: RPT-24-254N24.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPT-24. For BAC library availability, please contact Pierre de Jong (pjdejong@tigr.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end plate: 254 row: N column: 24  
Seq primer: T7  
Class: BAC ends.  
Location/Qualifiers

#### FEATURES

```

source
1. 749
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-254N24"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBACK1; Site_1: BamHI; Site_2: BamHI.
RPCI-24 Mouse BAC Library produced by Pletter de Jong. The
library was cloned in the pTARBACK1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT      216 a      170 c      165 g      198 t
ORIGIN

Query Match      19.0%, Score 36.4; DB 13; Length 749;
Best Local Similarity 54.5%; Pred. No. 0.39;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Oy 13 ggcacatgaatggtctgccaagcataatctgtctgtatcgtatcattcaactctcg 72
      |||||
Db 456 ggcatcttggtggcgttcacatgaacatcagctccgctgagatgattgctcattctgac 397
      |||||

Oy 73 gggaaatgagtgtgaaaccccttaaaacgggggttttcgagagaacatggcagatg 132
      |||||
Db 396 atgcatttggtgtgaaatccatgaatgacagtagccctccatttttgaaagccattgcttat 337
      |||||

Oy 133 catgatgacacaa 146
      |||||
Db 336 ctctgattgcaaaa 323

RESULT 2
Bg542199      852 bp.      mRNA      EST      03-APR-2001
LOCUS      602571637F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4696041 5',
DEFINITION      mRNA sequence.
ACCESSION      Bg542199
VERSION      Bg542199.1 GI:13534432
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 852)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LICM1522 row: j column: 10
High quality sequence stop: 365.
Location/Qualifiers
1. 852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4696041"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: lung; Vector: pDMR-LIB (Clontech); Site_1:
SfiI (ggcccctcgcgc); Site_2: SfiI (ggccattagcgc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCATTAATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGGCCGAGGGCGGCCGACATG-dT(30)BN-3' (where B = A,

```

c, or G and N - A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH/MCC Library."

BASE COUNT            215 a       260 c       199 g       178 t

ORIGIN

Query Match            18.0%; Score 34.6; DB 11; Length 852;  
Best Local Similarity    55.4%; Pred. No. 1.5;  
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy     38    aatctgtcgcattcgcttcatttcaaccttcggggaataagatgtaaacccctat 97  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     446    ACTATGCTCATTAATGGCCACAGCCACCAATFACATGGGGAACTTAAGTGCACGACCATAAT 505

Oy     98    aaacgcgggttttcgcagaacaatgcgctatgatcatgtgacaacatgactaaagca 157  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db     506    CAAGCGCGCGGTCTATCCTATATCGGACGAAGTCACATTTTTCATACATGAGACTACCCC 565

Oy     158 a 158  
|  
Db     566 A 566

RESULT     3  
A2964976/c

LOCUS       600 bp DNA                      GSS                      27-APR-2001

DEFINITION    ZM0234D2.R Mouse 10kb plasmid UNGC2M library Mus musculus genomic clone UNGC2M0234D2 R, DNA sequence.

ACCESSION    A2964976

VERSION      A2964976.1 GI:13836203

KEYWORDS     GSS.

SOURCE       house mouse.  
ORGANISM      Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murineae; Mus.

REFERENCE     1 (bases 1 to 600)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,R., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts  
Unpublished (2000)

JOURNAL       Contact: Robert B. Weiss  
COMMENT       University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA  
Tel.: 801 585 5606  
Fax: 801 585 7177  
Email: ddunne@genetics.utah.edu  
Insert length: 10000 . Std Error: 0.00  
plate: 0234 row: D column: 24  
Seq primer: CACACAGGAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 600.

TITLE          Location/Qualifiers  
1..600

FEATURES  
SOURCE

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UNG2M0234D24"  
/clone\_1lb="Mouse 10kb plasmid UNGC2M library"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, TI-resistant, F-"  
/note="Vector: PWD4nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces fragilis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia stipitis*, *Pichia guilliermondii*, *Pichia burtonii*, *Pichia kluyveri*, *Pichia guineensis*, *Pichia pastoris*, *Pichia anomala*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sobriophila*.

TITLE	Direct Submission
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : <a href="mailto:seqif@genoscope.cns.fr">seqif@genoscope.cns.fr</a> - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvatum</i> , <i>Saccharomyces</i>





Query Match	Best Local Similarity	17.3%	Score 33.2;	DB 13;	Length 630;				
Matches	74;	Conservative	0;	Mismatches	68;	Indels	0;	Gaps	0;
Oy	1	gctggcgcgcgccatgaagtgcttcgacagaatactctgtctgattcgtctatt	60						
Db	369	gCGAGCGCTGGCAGCGTGGGGCTGGCCGTCGCAATAAATCTCGCCGTGAAACAGTGCCTT	428						
Oy	61	tcaataaccttcggggaataatagatggaaaacctataaaagcggggttttgcagaac	120						
Db	429	TCCTAAAGACGTGTGTAATAAGTGAATATCTGTGTAAAGCGCCTGGCACATGCATATGG	488						
Oy	121	atgcgctagatcatcatgatgac	142						
Db	489	ACGCATATGTCCTCCATGGGAAAC	510						

RESULT	8
AZ931427	
LOCUS	AZ931427
DEFINITION	AZ931427 523 bp DNA GSS 01-APR-2001
ACCSSION	474.dhz5h10.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces
VERSION	AZ931427 unisporus genomic clone 474.dhz5h10.E1, DNA sequence.
KEYWORDS	AZ931427.1 GI:13502338
SOURCE	GSS.
ORGANISM	Saccharomyces unisporus.
	Saccharomyces unisporus.

REFERENCE	1 (bases 1 to 523)
AUTHORS	Cliffen,P.F., Hillier,L.W., Fulton,L., Graves,T., Miner,T., Gish
TITLE	M,R., Waterston,R.H. and Johnston,M.
JOURNAL	Surveying Saccharomyces genomes to identify functional elements by
COMMENT	Comparative DNA sequence analysis
	Unpublished (2001)
	Contact: Johnston M

Department of Genetics  
Washington University Medical School  
Box 8233, 4566 Scott Ave., St. Louis, MO 63110, USA  
Tel: 314 362 2735  
Fax: 314 362 7855  
Email: mjgenetics.wustl.edu  
Class: random plasmid subclone.  
Location/Qualifiers  
1. 523

```

/organism="Saccharomyces unisporus"
/strain="NRRL Y-1556 (CBS 398)"
/db_xref="taxon:27294"
/clone="474.dh265h10.s1"
/clone_lib="Saccharomyces unisporus NRRL Y-1556"
/note="Random genomic sequence"
BASE COUNT      176 a      90 c      86 g      170 t      1 others
ORIGIN

```

Query Match	17.2%	Score 33;	DB 13;	Length 523;
Best Local Similarity	54.5%;	Pred. No. 4.5;		
Matches 66;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;
21 agtggcttgacaagacataatcttgcgtgattcgtctatttcaatccttggggaata	80.			

Db	250	ATTAGTTTGAGAAAGTTTGATTCGGGCTCGACAGATCTAATTTGATACCGATTTGGGAGTA	309
Qy	81	gatygtgaaacccttataaaacgcgggttttcgcgaagaacatgcgcagtatcatlyatg	140
Db	310	AGAAAGCACCTCTGTTACCAAAAGATGCTCAATTAATGGCCATTGGGCGTTTCTACCTTG	369

QY	141	a	141
Db	370	A	370

RESULT	9
CNS04NR7	
LOCUS	957 bp
DEFINITION	DNA
	GSS
	21-MAY-2000
	Tetradon nigroviridis genome survey sequence T7 end of clone
	120161 of 1169016 from Tetradon nigroviridis

ACCESSION	AL299068
VERSION	AL299068.1
KEYWORDS	GS: genome survey sequence.
SOURCE	<i>Tetrahodon nigrovittatus</i> .
ORGANISM	<i>Tetrahodon nigrovittatus</i>

REFERENCE  
AUTHORS

Reest-Collins, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C. C.,  
Rost, M., & Williams, J. (2010). *Phylogenetic relationships of the  
Eukaryota: Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,  
Acanthopterygii, Neopterygii, Teleostei, Euteleostei, Neoteleostei,  
Acanthomorpha, Acanthopterygii, Percomorpha, Tetraodontiformes,  
Tetraodontidae, Tetraodon*.  
1 (bases 1 to 957)

TITLE	JOURNAL	REFERENCE
Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Unpublished	2 (bases 1 to 957)

AUTHORS  
 Boers-Cellarius, H., Jallion, O., Dasilva, C., Bouneau, J., Fisher, C.,  
 Bernot, A., Fizames, C., Wincker, P., Brotter, P., Queller, F.,  
 Saurin, W. and Weissenbach, J.  
 TITLE  
 Human gene number estimate provided by genome wide analysis using  
 retrodon microvirlidis DNA sequence  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Genomics  
 3 (phases 1 to 957)

**REVISIONS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Direct Submission to the EMBL/GenBank/DBJ databases  
Submitted (12-APR-2000)  
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

```

FEATURES
source
    location/Qualifiers
    1..957
    /organism="Tetradodon nigroviridis"
    /db_xref="taxon:99883"
    /clone="123G15"
    /clone_11b="G"
    /note="hemocope sequence ID : COBG123AD08LPI-end : T77
280 a 159 c 163 g 352 t 3 others
BASE COUNT
ORIGIN

```

Query Match	17.1%	Score 32.8	DB 13	Length 957
Best Local Similarity	54.0%	Pred. No. 5.9		
Matches 67	Conservative 0	Mismatches 57	Indels 0	Gaps 0

**Oy**

36 ataatcttgcgattcgtccatltttcaataaccticgggyaaatgatgtgaaaaacctt 95  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Dd**

229 ATTAGATATTCTGAAAGATTACTTGTATAAGCATGCCTTGCAAAAACAGCAAATACTTA 288

DQ 96 ataaacgcgggttttcgcagaacaatcgctagatcatgatgacacaatgactaag 15  
+  
Db 289 AGSAGGCGGTGATTTTGGCACCAAAAATCTTAATGTAAGAAGATGCACACTTTTCCTCG 340

|    |     |      |     |
|----|-----|------|-----|
| Qy | 156 | caaa | 159 |
|    |     |      |     |
| Db | 349 | AAAA | 352 |

RESULT 10  
BG062885

|            |            |             |         |                              |                    |
|------------|------------|-------------|---------|------------------------------|--------------------|
| LOCUS      | B6062885   | 552 bp      | mRNA    | EST                          | 25-JAN-2001        |
| DEFINITION | L0959C09-5 | NIA Mouse   | Newborn | Kidney cDNA Library2 (Short) | Mus                |
|            | musculus   | cdna        | clone   | L0959C09                     | 5', mRNA sequence. |
| ACCESSION  | B6062885   |             |         |                              |                    |
| VERSION    | B6062885.1 | GI:12533811 |         |                              |                    |





|                       |   |  |        |              |                       |
|-----------------------|---|--|--------|--------------|-----------------------|
| BASE COUNT            | 571 a   | 441 c  | 497 g  | 553 t        |                       |
| ORIGIN                |   |  |        |              | TEEVVQSLTMQNEVGSPC*   |
| Query Match           | 17.0%;  | Score 33.6;  | DB 12; | Length 2062; |                       |
| Best Local Similarity | 52.6%;  | Pred. No. 8;   |        |              |                       |
| Matches               | 71:   | Conservative   | 0:     | Mismatches   | 64; Indels 0; Gaps 0; |
| QY                    | 51  | tcgtcattttcaataaccttcgagggaataatagatgtgaaaacccctataaagcgcgggttt      | 110    |              |                       |
| Db                    | 1024  | tcgacctactcttccacgcttcacgctacgcttcacgcttcggaataattcgacaatgcactatttcc | 1083   |              |                       |
| QY                    | 111   | tcgcgaacacatgcgcctagatcatctatgtaacacatgagcactaaagcgaagtctgc          | 170    |              |                       |
| Db                    | 1084  | atgcagaaaagaccacttggatattacttactcacatcttacctgaagatcatcttcttcgg       | 1143   |              |                       |
| QY                    | 171   | cctgacccaagaag   | 185    |              |                       |
| Db                    | 1144  | cttcttcaagaggg   | 1158   |              |                       |
| RESULT 14             |   |  |        |              |                       |
| AK004888              |   |  |        |              |                       |
| LOCUS                 | 2366 bp   | mRNA   | HTC    | 05-JUL-2001  |                       |
| DEFINITION            | Mus musculus adult male liver cDNA, RIKEN full-length enriched library, clone:1300003007, full insert sequence.   |  |        |              |                       |
| ACCESSION             | AK004888  |  |        |              |                       |
| VERSION               | AK004888.1  | GI:12836413  |        |              |                       |
| KEYWORDS              | CAP trapper.  |  |        |              |                       |
| SOURCE                | Mus musculus (strain:C57BL/6J) adult male liver cDNA to mRNA, clone:1300003007.   |  |        |              |                       |
| ORGANISM              | Mus musculus  |  |        |              |                       |
| REFERENCE             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  |  |        |              |                       |
| AUTHORS               | 1 (bases 1 to 2366)   |  |        |              |                       |
| TITLE                 | Carninci, P. and Hayashizaki, Y.  |  |        |              |                       |
| JOURNAL               | High-efficiency full-length cDNA cloning  |  |        |              |                       |
| MEDLINE               | Methods in enzymology. 303, 19-44 (1999)  |  |        |              |                       |
| PUBMED                | 99279253  |  |        |              |                       |
| REFERENCE             | 10349636  |  |        |              |                       |
| AUTHORS               | 2 (bases 1 to 2366)   |  |        |              |                       |
| TITLE                 | Carninci, P., Shibata, Y., Hayatsu, N., Sugihara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.   |  |        |              |                       |
| JOURNAL               | Normalization and subraction of cap-trapper-selected cDNAs to   |  |        |              |                       |
| MEDLINE               | Prepare full-length cDNA libraries for rapid discovery of new genes   |  |        |              |                       |
| PUBMED                | Genome research. 10 (10), 1617-1630 (2000)  |  |        |              |                       |
| REFERENCE             | 20499374  |  |        |              |                       |
| AUTHORS               | 11042159  |  |        |              |                       |
| TITLE                 | 3 (bases 1 to 2366)   |  |        |              |                       |
| JOURNAL               | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M., Yoneda, T., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. |  |        |              |                       |
| TITLE                 | RIKEN integrated sequence analysis (RISA) system-384-format   |  |        |              |                       |
| JOURNAL               | sequencing pipeline with 384 multicapillary sequencer   |  |        |              |                       |
| MEDLINE               | Genome research. 10 (11), 1757-1771 (2000)  |  |        |              |                       |
| PUBMED                | 20530913  |  |        |              |                       |
| REFERENCE             | 11076861  |  |        |              |                       |
| AUTHORS               | 4 (bases 1 to 2366)   |  |        |              |                       |
| TITLE                 | The RIKEN Genome Exploration Research Group Phase II Team and the   |  |        |              |                       |
| JOURNAL               | FANTOM Consortium.  |  |        |              |                       |
| MEDLINE               | Functional annotation of a full-length mouse cDNA collection  |  |        |              |                       |
| PUBMED                | Nature 409, 685-690 (2001)  |  |        |              |                       |
| REFERENCE             | 5 (bases 1 to 2366)   |  |        |              |                       |
| AUTHORS               | Adachi, Y., Aizawa, K., Akahira, S., Akiura, T., Aono, H., Arai, A., Atakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M.,  |  |        |              |                       |

| FEATURES   | source   |
|------------|--|
| BASE COUNT | 609 a 353 c 604 g 598 t  |
| ORIGIN     | <p>Query Match 17.0% Score 32.6; DB 12; Length 2366;<br/>           Best Local Similarity 52.6%; Pred. No. 8.2;<br/>           Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;</p> <p>51 tcgtcatattcaactacctcggggaatagagtgaacaacctatataaagcgcggttt 110<br/>           Db 1328 TGGCGCTACTCTTCTTCTAGTTCACCCAGTACACCTCTCGAATAATTGCAACAAATCAGCATTTTC 1387<br/>           OY 111 tcgagaagaactgctcagtagtcatcttgatgcacaactatgaagcaaaagtctgtcc 170<br/>           Db 1388 ATGAGAGAAAGGCCATTGGATGATTCATCTACTCAGCATCTCAACCTGAAGATCATCTTCTTCGG 1447<br/>           OY 171 cctgaccacaagaag 185<br/>           Db 1448 CTCTTCAAGAGGCG 1462</p> |
| RESULT 15  |  |
| LOCUS      | AO787593 510 bp DNA GSS 03-AUG-1999  |
| DEFINITION | HS-2239_B1_F02_T7C Ctr Approved Human Genomic Sperm Library D Homo   |
| ACCESSION  | AO787593   |
| VERSION    | AO787593.1 GI:5695217  |
| KEYWORDS   | GSS.   |
| SOURCE     | human.   |

| ORGANISM              | Homo sapiens  |
|-----------------------|---|
| Eukaryota:            | Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi;  |
| Mammalia:             | Eutheria; Primates; Catarrhini; Homnidae; Homo.   |
| REFERENCE             | 1 (bases 1 to 510)  |
| AUTHORS               | Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T.,<br>Keller,A., Shaker,R., Furlong,J.J., Young,J.J., Zhao,S., Adams,M.D. and<br>Hood,L.  |
| TITLE                 | Sequence-tagged connectors: A sequence approach to mapping and<br>scanning the human genome   |
| JOURNAL               | Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)   |
| MEDLINE               | 99380589  |
| COMMENT               | Contact: Mahairas GG, Wallace JC, Hood L<br>High Throughput Sequencing Center<br>University of Washington<br>401 Queen Anne Avenue North, Seattle, WA 98109, USA<br>Tel: (206) 616-3618<br>Fax: (206) 616-3887<br>Email: jwallace@u.washington.edu<br>Clones may be purchased from Research Genetics (info@resgen.com).<br>BAC end Web Server: http://www.htsc.washington.edu<br>Plate: 2239 row: L column: 3<br>Seq primer: T7<br>Class: BAC ends<br>High quality sequence stop: 510.<br><br>Location/Qualifiers<br>1..510<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clo_nref="plate:2239 Col=3 Row=L"<br>/clone_1lb="CIT Approved Human Genomic Sperm Library D"<br>/sex="male"<br>/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in<br>E-Coli DH10B" |
| BASE COUNT            | 156 a     104 c     90 g     157 t     3 others   |
| ORIGIN                |   |
| Query Match           | 16.9%; Score 32.4; DB 13; Length 510;   |
| Best Local Similarity | 52.2%; Pred. No. 6.9;   |
| Matches               | 72; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  |
| OY                    | 25 gccgtacgaagcataactctgctcgtatcgctcatatttcaaacaccttcgggaaatagatg 84<br>  |
| Db                    | 257 GCTTAAGAAATATAAGGTCACAATAATCATCGAGACTGTGGTTCCTTTCAACAATAATTATC 316  |
| OY                    | 85 tgaaaaccctataaacygggttttcgcagaacaacatgcyctgatcatatgtagaca 144<br>  |
| Db                    | 317 ATCTTAAGATCATATAAACAGGCCTTCTCTCACAAATGTCATCCATCATGTTTGTCAT 376  |
| OY                    | 145 catgactaagcaaaagt 162<br>   |
| Db                    | 377 GATTGTATATATAAAAAGT 394   |

Search completed: March 28, 2002, 14:42:22  
Job time: 7576 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:16:11 ; Search time 1551.88 Seconds  
(without alignments)  
1371.328 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_472\_600

Perfect score: 129  
Sequence: 1 tgcattctgctgcgtacgca.....tattccgagactagttcagcg 129

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_lm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_scs:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcgo\_hum:\*  
31: em\_hcgo\_inv:\*  
32: em\_hcgo\_rtd:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID         | Description            |
|------------|-------|-------------|--------|---------------|------------------------|
| 1          | 31    | 24.0        | 5832   | 1 AF105113    | AF105113 Streptococcus |
| 2          | 31    | 24.0        | 18754  | 1 AF094575    | AF094575 Streptococcus |
| 3          | 29.6  | 22.9        | 174118 | 2 AC024405    | AC024405 Homo sapi     |
| 4          | 29.6  | 22.9        | 185341 | 2 AC069403    | AC069403 Homo sapi     |
| 5          | 29.6  | 22.9        | 331109 | 2 AL162413    | AL162413 Homo sapi     |
| 6          | 29.4  | 22.8        | 179759 | 2 AC092021    | AC092021 Homo sapi     |
| 7          | 29.4  | 22.8        | 190919 | 2 AC091731    | AC091731 Homo sapi     |
| 8          | 29.4  | 22.8        | 202224 | 2 AL356219    | AL356219 Homo sapi     |
| 9          | 29.4  | 22.8        | 203674 | 2 AC025669    | AC025669 Mus muscu     |
| 10         | 29.4  | 22.8        | 230838 | 2 AL591946    | AL591946 Mus muscu     |
| 11         | 29.2  | 22.6        | 1877   | 10 AF176007   | AF176007 Mus muscu     |
| 12         | 29.2  | 22.6        | 1928   | 10 AF192558   | AF192558 Mus muscu     |
| 13         | 29.2  | 22.6        | 2193   | 10 RL044387   | RL044387 Rhizobium     |
| 14         | 29.2  | 22.6        | 5263   | 1 RLRES       | RLRES Rhizobium        |
| 15         | 29.2  | 22.6        | 21984  | 10 AB040292   | AB040292 Mus muscu     |
| 16         | 29    | 22.5        | 138041 | 9 CDS01D01    | AL133245 BAC seque     |
| 17         | 29    | 22.5        | 16318  | 9 CDS01D06    | AL133243 BAC seque     |
| 18         | 29    | 22.5        | 189008 | 2 AP001491    | AP001491 Homo sapi     |
| 19         | 29    | 22.5        | 201657 | 2 AL596187    | AL596187 Mus muscu     |
| 20         | 29    | 22.5        | 232190 | 2 AL596025    | AL596025 Mus muscu     |
| 21         | 28.8  | 22.3        | 35694  | 8 SP8C23G7    | AL035065 S.pombe       |
| 22         | 28.8  | 22.3        | 152148 | 2 AC063952    | AC063952 Homo sapi     |
| 23         | 28.8  | 22.3        | 185401 | 2 AC024927    | AC024927 Homo sapi     |
| 24         | 28.6  | 22.2        | 278    | 1 BOV279043   | BOV279043 Bacteroid    |
| 25         | 28.6  | 22.2        | 10353  | 1 AE004965    | AE004965 Pseudomon     |
| 26         | 28.6  | 22.2        | 110000 | 2 LMFCHR16_02 | Continuation (3 of     |
| 27         | 28.6  | 22.2        | 171490 | 2 AL356092    | AL356092 Homo sapi     |
| 28         | 28.6  | 22.2        | 180292 | 9 AL359269    | AL359269 Human DNA     |
| 29         | 28.6  | 22.2        | 201320 | 2 AC073816    | AC073816 Mus muscu     |
| 30         | 28.6  | 22.2        | 247196 | 2 AC073822    | AC073822 Mus muscu     |
| 31         | 28.4  | 22.0        | 731    | 8 ZMHMGD1     | Y08807 Z.mays mRNA     |
| 32         | 28.4  | 22.0        | 13104  | 1 AE000985    | AE000985 Archaeogl     |
| 33         | 28.4  | 22.0        | 65762  | 2 AC079913    | AC079913 Homo sapi     |
| 34         | 28.2  | 21.9        | 173515 | 2 AC073394    | AC073394 Homo sapi     |
| 35         | 28.2  | 21.9        | 177642 | 2 HSDJ90K10   | AL122034 Homo sapi     |
| 36         | 28.2  | 21.7        | 332635 | 1 AP003005    | AP003005 Mesophilic    |
| 37         | 28    | 21.7        | 2384   | 1 SRMSIK      | Y08921 S.reliculi      |
| 38         | 28    | 21.7        | 143515 | 8 AP002526    | AP002526 Oryza sat     |
| 39         | 27.8  | 21.6        | 110019 | 9 AC068860    | AC068860 Homo sapi     |
| 40         | 27.8  | 21.6        | 145428 | 2 AC026068    | AC026068 Homo sapi     |
| 41         | 27.8  | 21.6        | 147484 | 2 AL590088    | AL590088 Homo sapi     |
| 42         | 27.8  | 21.6        | 203668 | 2 AL592148    | AL592148 Homo sapi     |
| 43         | 27.6  | 21.4        | 2150   | 1 AF017790    | AF017790 Homo sapi     |
| 44         | 27.6  | 21.4        | 8280   | 1 PSEPLAS     | X66604 Pseudomonas     |
| 45         | 27.6  | 21.4        | 9617   | 1 RLX12758    | RLX12758 R.leguminos   |

## ALIGNMENTS

RESULT 1  
LOCUS AF105113 5832 bp DNA BCT 09-SEP-1999  
DEFINITION Streptococcus pneumoniae type 19A putative oligosaccharide repeat unit transporter (cps19A) gene, partial cds; UDP-N-acetyl glucosamine-2-epimerase (cps19AK), glucose-1-phosphate thymidyl transferase (cps19AL), dTDP-4-keto-6-deoxyglucose-3,5-epimerase (cps19AM), dTDP-glucose-4,6-dehydratase (cps19AN), and dTDP-L-rhamnose synthase (cps19AO) genes, complete cds; and AL1A (aIIA) gene, partial cds.  
ACCESSION AF105113  
VERSION AF105113.1 GI:4406246  
KEYWORDS

SOURCE Streptococcus pneumoniae.  
ORGANISM Streptococcus pneumoniae  
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.  
REFERENCE 1 (bases 1 to 5832)  
AUTHORS Morona,J.K., Morona,R. and Paton,J.C.  
TITLE Comparative genetics of capsular polysaccharide biosynthesis in





|           |  |
|-----------|--|
| ACCESSION | AF094575   |
| VERSION   | AF094575.1   |
| KEYWORDS  | GI:3907597   |
| SOURCE    | Streptococcus pneumoniae   |
| ORGANISM  | Streptococcus pneumoniae   |
| REFERENCE | Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.   |
| AUTHORS   | 1 (bases 1 to 18754)   |
| TITLE     | Morona, J. K., Morona, R. and Paton, J. C.   |
| JOURNAL   | Analysis of the 5' portion of the type 19a capsular locus identifies two classes of cpsC, cpsD, and cpsE genes in Streptococcus pneumoniae   |
| MEDLINE   | J. Bacteriol. 181 (11), 3599-3605 (1999)   |
| PUBMED    | 99287847   |
| REFERENCE | 2 (bases 1 to 18754)   |
| AUTHORS   | Morona, J. K., Morona, R. and Paton, J. C.   |
| TITLE     | Comparative genetics of capsular polysaccharide biosynthesis in Streptococcus pneumoniae types belonging to serogroup 19   |
| JOURNAL   | J. Bacteriol. 181 (17), 5355-5364 (1999)   |
| MEDLINE   | 99395043   |
| REFERENCE | 3 (bases 1 to 18754)   |
| AUTHORS   | Morona, J. K., Morona, R. and Paton, J. C.   |
| TITLE     | Submitted (23-SEP-1998) Microbial Microbiology Unit, Women's and Children's Hospital, King William Rd, North Adelaide, SA 5006, Australia  |
| FEATURES  | Location/Qualifiers  |
| SOURCE    | 1..18754   |
| gene      | /organism="Streptococcus pneumoniae"   |
| CDS       | /serotype="19a"  |
| gene      | /db_xref="taxon:1313"  |
| CDS       | <1..9  |
| gene      | /gene="dexB"   |
| CDS       | <1..9  |
| gene      | /gene="dexB"   |
| CDS       | /codon_start=1   |
| gene      | /transl_table=11   |
| CDS       | /product="DexB"  |
| gene      | /protein_id="AAC78662.1"   |
| CDS       | /db_xref="GI:3907598"  |
| gene      | /translation="VE"  |
| CDS       | complement(161..1647)  |
| gene      | /note="Insertion sequence; 3' truncated"   |
| CDS       | /rpt_family="IS1202"   |
| gene      | /rpt_type="dispersed"  |
| CDS       | 1733..3178   |
| gene      | /gene="cps19a"   |
| CDS       | 1733..16682  |
| gene      | /gene="capsular polysaccharide biosynthesis operon"  |
| CDS       | 1733..3178   |
| gene      | /gene="cps19a"   |
| CDS       | /codon_start=1   |
| gene      | /transl_table=11   |
| CDS       | /product="putative regulatory protein Cps19a"  |
| gene      | /protein_id="AAC78663.1"   |
| CDS       | /db_xref="GI:3907599"  |
| gene      | /translation="MSRRFKRSQKRVSVNIVLLTYLLVCFLLFLFTYNIILAF RYLNLYVPLVLYLVALLLYKKAEFTIPLVPSILVSSYSLFAVOOFGILNR LNATNSYSESVAVLADSEINVOQLMPTGDNENIOKLLADIKSSONTDT VNGSSSLAYKXSIAGETKATVINSFENIIESEHPDYASKIKITTKKPTKVEAP KISKDSFNIVSGIDTYGPIISSVRSYVNIILMTVNDTKILLTTTPRAVYPIADG GNNOKDKLTHAGIYGVDSIHTLENIYGVINTYVRLNFTSFLKIDLLGCVVYNDG DFTSLNGKFFHPYGVNHLDEQALGFERYERSLADGHDGRNQOKVIAIILQKITS EALKNSRIIDSDSIOTNMPLETMINLVNAQESGCTYKVSQDLKGRGRDLPSEY AMPDSNLYMEINDSSLASVKTAIQDVLBGR" |
| gene      | 3180..3911   |
| CDS       | /gene="cps19a"   |
| gene      | /gene="cps19a"   |
| CDS       | 3180..3911   |
| gene      | /gene="cps19a"   |
| CDS       | /codon_start=1   |
| gene      | /transl_table=11   |
| CDS       | /product="Cps19a"  |
| gene      | /protein_id="AAC78664.1"   |
| CDS       | /db_xref="GI:3907600"  |
| gene      | /translation="MIDHSHIVDVDDGPKRSRESKALLAESYRQGVRTVTSNHR KQMEPEEKIAENFLOVREIAKVASDLYVAAEYYSVDVLDKEKRIPYPLANS RYALIERSNMTPYRDIHSALSKITLMTITPVIAIEEDYDLNENKREILINNGCYT QVSSHAVLKKLREGRYKFMKRAQYLELDVHLIASDMHNDSPRHMAEADLVAS OKYGEAKOELFDINPKRIYQDLI"   |
| gene      | 3917..4612   |
| CDS       | /gene="cps19a"   |
| gene      | 3917..4612   |
| CDS       | /gene="cps19a"   |
| gene      | /function="chain length regulation and export"   |
| CDS       | /codon_start=1   |
| gene      | /transl_table=11   |
| CDS       | /product="Cps19a"  |
| gene      | /protein_id="AAC78665.1"   |
| CDS       | /db_xref="GI:3907601"  |
| gene      | /translation="MKKEONTIETDFQILTKMKLILLVALVYAGAAVSTFI VKPEYISTRIYVVRNQGDKPGLTNDQAGYLVADYEIILSDVLEKRVATNKL DMPAKLTISKVQVTPADTRIVSISVDKQPEEASRIANSLREVAEKIYATVRSVY TTLEEARPATPSPSNVRRNTLVGLGAAVYTYVLLIELDTRVRRPDEVQVLM PLGVDPDFDKMK"   |
| gene      | 4622..5311   |
| CDS       | /gene="cps19a"   |
| gene      | 4622..5311   |
| CDS       | /gene="cps19a"   |
| gene      | /function="chain length regulation and export"   |
| CDS       | /codon_start=1   |
| gene      | /transl_table=11   |
| CDS       | /product="Cps19a"  |
| gene      | /protein_id="AAC78666.1"   |
| CDS       | /db_xref="GI:3907602"  |
| gene      | /translation="MPTLEISOAKLELVKKAEEYNAICTNLOSGDDLKVFSTSVK AGEGKTTSTNIMAAEACAGYKTLIDADNRNMSGVCFRSRRTIGTFEELSGTSG SOCIQCTNENLFIQAGSVSPNPTALDSKNTTMEITRKRFDYVYDTRAGVYI DAATIMOKCDASLIVTKAGETKRDLQAKOEQYKSCVGLVNLKDLTSVERKGSY GSYGNRKOK"   |
| gene      | 5326..6687   |
| CDS       | /gene="cps19a"   |
| gene      | 5326..6687   |
| CDS       | /gene="cps19a"   |
| gene      | /codon_start=1   |
| CDS       | /transl_table=11   |
| gene      | /product="glucosyl-1-phosphate transferase Cps19a"   |
| CDS       | /protein_id="AAC78667.1"   |
| gene      | /db_xref="GI:3907603"  |
| CDS       | /translation="MSEKLAQPSVAIVOSFLVLLTYLLASVRETEIVSTAIVYL HYFVYISDGRNPFKRGYILEVQTLKILFPALASISNPFLEDRFSISRGMITF LLHVELVLYLVNPFIRKMYMRKAVPNFRGSKILLTATSVEKVLRLIESDVDEL VAVSYLDKDFQHDLDLVVAEGEIVNEATHEVDEVEINLPSKYNIGELVSQFETMG IDVTVINLADPDMARNRKOICEMAGLNVTTFSTYKTSYVAKRIIDVSLVGLILG LVSIVLVPILIRKDGSAFQOTRIGKGRHFTFYKFRSMCVDAEKKRELMQONTMOG GMFKVDDDPRIITKIGHPIRTKSIDLPQFNVLKGMKSYCTGPPYDEVTEHTPECK RRLSFKPGITGLMOVSGRSEIKNFDEVKLDVAVIDGWTIMKIDIELTKVAVLMD GAK"               |
| gene      | 6693..7436   |
| CDS       | /gene="cps19a"   |
| gene      | 6693..7436   |
| CDS       | /gene="cps19a"   |
| gene      | /codon_start=1   |
| CDS       | /transl_table=11   |
| gene      | /product="putative N-acetyl-mannosamine transferase Cps19a"  |
| CDS       | /protein_id="AAC78668.1"   |
| gene      | /db_xref="GI:3907604"  |
| CDS       | /translation="ANERIOIDGITDPLMKFTVDAVEQYVLKKHPLHMGVNAKTI MOCYEDKELIKIVNESGIIINADGASVLAASKLGTVPPEVAGIDILMKHLLELSNKG YSVYFPGAKEDVLDQMKIIFERYPALNVGVNGYFSPEDEKHIEDIKERKPDVF VGTISPEKKEYIIOSFMDNGINAVFMGVGGSFVLSGHIKRAPLMQNAHLEMTFRVAN EPRKLFKRYFVGVSFLIRKVLRAKRGYKL"   |
| gene      | 7475..8257   |
| CDS       | /gene="cps19a"   |
| gene      | 7475..8257   |
| CDS       | /gene="cps19a"   |

```

/gene="cps19a"
/function="putative protein involved in formation of
phospho-diester linkage"
/codon_start=1
/transl_table=11
/product="Cps19a"
/protein_id="AAC78669.1"
/db_xref="GI:3907605"
/translation="MDAIEKQKICKENNDIFELGSGVLAQVYKDFIPMDMDIA
VREGYKPGIEKDRITAGKYOVLAQYCDPLHCYFPLLEDERKRGCPRTNL
GHLHDITPLDAPNHSFKRTKFGKVTYRRLASGTYVDHDMHAKKLLTIGF
FKKIDFAKLEPQNVYRRLDNLKRYDMKQYAGTINSLAKEMPEVIEGEEVK
PEDAFFKVPTEYDRYKRLXGENTLHEPSDEKSHLGCG"
8259. 9137
/gene="cps19a"
8259. 9137
/gene="cps19a"
/codon_start=1
/transl_table=11
/product="putative rhamnosyl transferase Cps19a"
/protein_id="AAC78670.1"
/db_xref="GI:3907606"
/translation="MCTYILHLYKYLETSCVKSIRKGNSTKQIIVINDSNNGTG
EKIQELYSDESIDVLIHNNAGFARGNNAVQAFKEXKPPFVIMNDIEETEEF
EKIVDIYRKEKFLGPDIFSTYQLOHNPRLHYEYKALNEKKGSOVLA
LKIKCMKSSKVLRTAIONRRKRSKQVENDILHGSFTVSRDLKEEYAFN
PNTFYETELDYAEELKGYKRIYTPRKIVLHONVATNOYTNLVENTLESKNCNF
ESTSYFLMKRNEGV"
9138. 10472
/gene="cps19a"
9138. 10472
/gene="cps19a"
/codon_start=1
/transl_table=11
/product="polysaccharide polymerase Cps19a"
/protein_id="AAC78671.1"
/db_xref="GI:3907607"
/translation="MTEYELGLTLELFFEFAPNODLIAPVVMVAFLISSVFA
LIYNQNMIEKSGAVYLLISGLIIVESMPLALNSPLNTKIKVDRLIDIOKMLIA
TIYDILITLYRREIHNALSHGTSNCFQFPFNATSYBEELVIRISIRLITID
VSAIFGTFFINFTYSHKRSKDLILPLIFLISKLLSGSLDITKILAVYMA
YIQQRKVGMDVISHKRYMRLEFGLINGIPFYSLFSGSTRFVESISTYLG
SIOHNOYIONIGVAVFEDSEFVAIMNLGIVYNSVHLEFQIGTMCNV
TEFRPMDFGIVAGVIFSEYVGFALFYKLKRSRAGFKDITITISYFYVIFL
SSIOYSEFTTISLFLVFLVLYLMAFYMNDFHRKLVITLSDTSIKSEN"
10497. 11924
/gene="cps19a"
10497. 11924
/gene="cps19a"
/codon_start=1
/codon__start=1

Query Match 24.0%; Score 31. DB 1; Length 18754;
Best Local Similarity 59.8%; Pred. No. 17;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Oy 32 attactgaacaaagaccgcagagatgctgtctttgtgtgctgtga 91
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13828 ATTACGAGCTTAACAAGCTTAACCTTAATCGGGGACCTTCTGTGAGCTGATGGG 13887

Oy 92 cgtgtgttcacacgcattatattccgga 118
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 13888 CCTGCTTTGCTGCTGATGATACGGCA 13914

RESULT 3
AC024405/174118 bp DNA HTG 11-AUG-2001
LOCUS Homo sapiens chromosome 11 clone RP11-574M7 map 11, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
AC024405
AC024405.6 GI:15148297
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 174118)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 11, clone RP11-574M7
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 174118)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,M., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuirk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 11, 2001 this sequence version replaced g1:15144553.
All repeats were identified using RepeatMasker.
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 574_M7
Center clone name: 574_M7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 174118: contig of 174118 bp in length.
Location/Qualifiers
1. 174118
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-574M7"
/clone.lib="RP11-574M7 Human Male BAC"
BASE COUNT 54806 a 34652 c 31384 g 53276 t
ORIGIN
Query Match 22.9%; Score 29.6; DB 2; Length 174118;
Best Local Similarity 59.5%; Pred. No. 53;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Oy 25 ttgcagcattacgaacaaagaccgcagagatgctgtctttgtgtgtgt 84
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 76069 TTGGTCCAGTCCCTGACCAAGTCTTATCCAGATCTGATGCTCTTGTCCAGTTGGTCTGCT 76010
OY 85 gctgtgacgtgtgtccaccgcta 108
      |||| 11 ||||| 111
Db 76009 CCGTCCAGTCCCTGCTCCAGTCTTA 75986

RESULT 4
AC069403 185341 bp DNA HTG 22-FEB-2001
LOCUS Homo sapiens chromosome 11 clone RP11-54B5, WORKING DRAFT SEQUENCE,
DEFINITION 2 unordered pieces.
AC069403
AC069403.11 GI:11878497
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 185341)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Fiederspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmly,J.,
Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 185341)
Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Fiederspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmly,J.,
Yu,S. and Davis,R.W.
Unpublished

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Dec 18, 2000 this sequence version replaced gt:11693354.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDRSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: RP11-54B5
Center clone name: RP11-54B5
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 100% of reads
Sequencing Vector: plasmid; accession: 0% of reads
Chemistry: Dye-terminator; 1% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184714 bases at least Q40
Consensus quality: 184960 bases at least Q30
Consensus quality: 185096 bases at least Q20
Insert size: 113180; agarose-fp
Insert size: 185241; sum-of-contigs
Quality coverage: 15.2x in Q20 bases; agarose-fp
Quality coverage: 9.3x in Q20 bases; sum-of-contigs.
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 43623: contig of 43623 bp in length
* 43624 43723: gap of unknown length
* 43724 185341: contig of 141618 bp in length.
Location/Qualifiers
1. 185341
/organism="Homo sapiens"
/db_xref="taxon:9606"

Db 71459 TTGGTCCAGTCCCTGACCAAGTCTTATCCAGATCTGATGCTCTTGTCCAGTTGGTCTGCT 71518
OY 85 gctgtgacgtgtgtccaccgcta 108
      |||| 11 ||||| 111
Db 71519 CCGTCCAGTCCCTGCTCCAGTCTTA 71542

RESULT 5
AL162413/c
LOCUS Homo sapiens chromosome 9 clone RP11-109M17, *** SEQUENCING IN
DEFINITION PROGRESS ***; 11 unordered pieces.
AC069403
AC069403.14 GI:15131970
VERSION HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 331109)
AUTHORS Corby,N.
TITLE Direct Submission
JOURNAL Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Aug 9, 2001 this sequence version replaced gt:15029110.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA109M17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 326784 bases at least Q40
Consensus quality: 328297 bases at least Q40
Consensus quality: 329246 bases at least Q20
Insert size: 330109; sum-of-contigs
Insert size: 303557; 19.2% error; agarose-fp
Quality coverage: 6.30x in Q20 bases; sum-of-contigs quality
Coverage: 7.15x in Q20 bases; agarose-fp
NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 59586: contig of 59586 bp in length

```

```

* 59587 59686: gap of 100 bp
* 59687 158058: contig of 98372 bp in length
* 158059 158158: gap of 100 bp
* 158159 265019: contig of 106861 bp in length
* 265020 265119: gap of 100 bp
* 265120 282213: contig of 17094 bp in length
* 282214 282313: gap of 100 bp
* 282314 288703: contig of 6390 bp in length
* 288704 288803: gap of 100 bp
* 288804 309775: contig of 20972 bp in length
* 309776 309875: gap of 100 bp
* 309876 312053: contig of 2178 bp in length
* 312054 318992: contig of 6839 bp in length
* 318993 319092: gap of 100 bp
* 319093 322077: contig of 2985 bp in length
* 322078 322177: gap of 100 bp
* 322178 328201: contig of 5924 bp in length
* 328202 331109: contig of 2908 bp in length.
* 328202 Location/Qualifiers
FEATURES
SOURCE
1..331109
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-109M17"
/clone.lib="RPC1-11.1"
1..59586
/feature="assembly_fragment:01096"
fragment_chain:1
misc_feature
59687..158058
/feature="assembly_fragment:00965"
fragment_chain:1
misc_feature
158159..265019
/feature="assembly_fragment:03121"
fragment_chain:2
misc_feature
265120..282213
/feature="assembly_fragment:04814"
fragment_chain:2
misc_feature
282314..288703
/feature="assembly_fragment:03747"
fragment_chain:3
misc_feature
288804..309775
/feature="assembly_fragment:00694"
fragment_chain:3
misc_feature
309876..312053
/feature="assembly_fragment:00673"
312154..318992
/feature="assembly_fragment:01842"
misc_feature
319093..322077
/feature="assembly_fragment:03292"
misc_feature
322178..328101
/feature="assembly_fragment:03601"
misc_feature
328202..331109
/feature="assembly_fragment:04963"
BASE COUNT 105642 a 64007 c 60342 g 100117 t 1001 others
ORIGIN

```

```

Query Match 22.9% Score 29.6 DB 2 Length 331109;
Best Local Similarity 59.3% Pred. No. 54;
Matches 50: Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

OY 25 ttccagcattcagcacaagcccgaccgagatgtcgggtctttgtgtgt 84
Db 174067 TTGTCACAGTCTTACACAGTCTTATCCAGATCTGCATGCTTGTGTCAGTTGTGCTGCT 174008
OY 85 gctgtgacgtgtgtgtccacacgta 108
Db 174007 CCTGTCACAGTCTTCCACAGTCTTA 173984

```

RESULT 6

```

AC092021/c
LOCUS AC092021 179759 bp DNA HTG 12-AUG-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-420H12, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
ACCESSION AC092021
VERSION AC092021.2 GI:15148366
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179759)
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL 2 (bases 1 to 179759)
REFERENCE
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2001 this sequence version replaced gi:14423621.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0420H12
----- Summary Statistics -----
Sequencing vector: M13: 5%
Sequencing vector: plasmid: 95%
Chemistry: Dye-terminator ET; 1% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165447 bases at least Q40
Consensus quality: 167029 bases at least Q30
Consensus quality: 178591 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 10000: contig of 10000 bp in length
* 10001 10100: gap of unknown length
* 10101 11203: contig of 1103 bp in length
* 11204 11303: gap of unknown length
* 11304 21454: contig of 10151 bp in length
* 21455 21554: gap of unknown length
* 21555 44442: contig of 22888 bp in length
* 44443 44542: gap of unknown length
* 44543 83511: contig of 38969 bp in length
* 83512 83611: gap of unknown length
* 83612 178527: contig of 94916 bp in length
* 178528 178627: gap of unknown length
* 178628 179759: contig of 1132 bp in length.
Location/Qualifiers
1..179759
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-420H12"
1..10000
/feature="assembly_name:Contig1"
10101..11203
/feature="assembly_name:Contig14"
11304..21454
/feature="assembly_name:Contig15"

```

```

FEATURES
SOURCE
misc_feature
1..10000
/feature="assembly_name:Contig1"
10101..11203
/feature="assembly_name:Contig14"
11304..21454
/feature="assembly_name:Contig15"

```

```

misc_feature      21555..44442
                  /note="assembly_name:Contig16
                  clone_end:SP6
                  vector_side:left"
misc_feature      44343..83511
                  /note="assembly_name:Contig17
                  clone_end:T7
                  vector_side:left"
misc_feature      83612..178527
                  /note="assembly_name:Contig18"
                  /note="assembly_name:Contig19"
                  /note="assembly_name:Contig20"
BASE COUNT      49208 a 40863 c 39822 g 49254 t 612 others
ORIGIN
Query Match      22.8%; Score 29.4; DB 2; Length 179759;
Best Local Similarity 56.8%; Pred. No. 62;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
OY 24 ctctgagcattactgcagcaagaagaccgacgagatgctggggtcttcttggtg 83
Db 30322 CCTCCTGAGTGCGCTGACATTTACAGCACCACACAGATGCTCGCTAATTTGTTATTT 30263
OY 84 tgcgtgacgtgtgttccaacgcgtattatccgga 118
Db 30262 TTAGTAGAGATGGGCTTTCACCATGTTAGCCAGGA 30228

RESULT 7
AC091731 190919 bp DNA HTG 05-JUL-2001
LOCUS      Homo sapiens chromosome UNK clone RP11-343F24, WORKING DRAFT
DEFINITION
AC091731
AC091731.2 GI:14572169
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 190919)
AUTHORS      Waterston, R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 190919)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (25-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Jun 27, 2001 this sequence version replaced gi:14196415.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: HNH0343F24

----- Summary Statistics -----
Sequencing vector: M13;
Chemistry: Dye-Primer ET; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180703 bases at least Q40
Consensus quality: 186902 bases at least Q30
Insert size: 203000; agarose-fp
Quality coverage: 4.39 in Q20 bases; agarose-fp
Quality coverage: 4.73 in Q20 bases; sum-of-coverage

```

```

FEATURES
SOURCE
1..190919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-343F24"
1..18062
/note="assembly_name:Contig10"
18163..41855
/note="assembly_name:Contig11"
41956..74273
/note="assembly_name:Contig12"
74376..150879
/note="assembly_name:Contig13"
150980..152042
/note="assembly_name:Contig14"
152143..154184
/note="assembly_name:Contig15"
154285..157461
/note="assembly_name:Contig16"
157562..162983
/note="assembly_name:Contig17"
163084..166817
/note="assembly_name:Contig18"
166918..171135
/note="assembly_name:Contig19"
171236..190919
/note="assembly_name:Contig20"
BASE COUNT      51253 a 42453 c 43073 g 53129 t 1011 others
ORIGIN
Query Match      22.8%; Score 29.4; DB 2; Length 190919;
Best Local Similarity 56.8%; Pred. No. 62;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
OY 24 ctctgagcattactgcagcaagaagaccgacgagatgctggggtcttcttggtg 83
Db 142084 CCTCCTGAGTGCGCTGACATTTACAGCACCACACAGATGCTCGCTAATTTGTTATTT 142025
OY 84 tgcgtgacgtgtgttccaacgcgtattatccgga 118
Db 142024 TTAGTAGAGATGGGCTTTCACCATGTTAGCCAGGA 141990

```

```

RESULT      9
LOCUS       AL356219      28-APR-2001
DEFINITION  Homo sapiens chromosome 9 clone RP11-430C15, *** SEQUENCING IN
ACCESSION   AL356219
VERSION     AL356219.11  GI:13445372
KEYWORDS    HOMO
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 202224)
AUTHORS     Lloyd, D.
TITLE       Direct Submission
JOURNAL     Submitted (27-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requesters: clonerequest@sanger.ac.uk
On Mar 24, 2001 this sequence version replaced gi:13396524.

COMMENT     ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba430C15
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 202171 bases at least Q40
Consensus quality: 202224 bases at least Q30
Consensus quality: 202224 bases at least Q20
Insert size: 202224; sum-of-contigs
Insert size: 181975; 8.7% error; agarose-fp
Quality coverage: 10.85x in Q20 bases; sum-of-contigs Quality
coverage: 12.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES             location/qualifiers
     source           1..202224
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-430C15"
                     /clone_1fb="RPC1-11.2"
                     /note="202224
                     /note="assembly-fragment:04639
                     /clone_end:SP6
                     vector_side:left"
     misc_feature     59963 a 39332 c 40682 g 62247 t
     ORIGIN
BASE COUNT      59963 a 39332 c 40682 g 62247 t
Query Match      22.8%; Score 29.4; DB 2; Length 202224;
Best Local Similarity 60.8%; Pred. No. 62;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

42 acaagaccgcagcagagatgctggcgctctttgtgtgtgtgtgtgacgtgtgttc 101
| 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 | 111 |
Db 180928 AGAAACTCACCACCTTGACACAAAGTGAACCTCTTGTGTGTGTGTGTGTGCA 180987
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
QY 102 aacgtatattccgcgact 120
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
Db 180988 GAGGTGTGTTTCCGGCT 181006
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

```

```

LOCUS      AC025669
DEFINITION Mus musculus chromosome 14 clone RP23-5016 map 14, *** SEQUENCING
IN PROGRESS ***, 5 ordered pieces.
ACCESSION  AC025669
VERSION    AC025669.8 GI:15028574
KEYWORDS   HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE     Mus musculus
ORGANISM   house mouse.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 203674)
REFERENCE  Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Mus musculus chromosome 14, clone RP23-5016
REFERENCE  Unpublished
AUTHORS     2 (bases 1 to 203674)
           Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
           Anderson, S., Baldwin, J., Barra, N., Bastien, V., Beda, F.,
           Boguslavsky, I., Boukhalter, B., Brown, A., Burkett, G.,
           Campolongo, A., Castle, A., Chepel, Y., Colangelo, M., Collins, S.,
           Collamore, A., Cooke, P., Dearlouis, K., Dewar, K., Diaz, J.S.,
           Dodge, S., Domino, M., Doyle, M., Ferrera, P., Fitzhugh, W., Gage, D.,
           Galagan, J., Gardina, S., Ginde, S., Coyette, M., Graham, L.,
           Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Hottel, L.,
           Hovland, D.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
           Klein, J., LaCoque, K., Lamazares, R., Landers, T., Lehoccky, J.,
           Levine, R., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
           McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R.,
           Melrim, J., Menus, L., Mihov, T., Miranda, C., Mlangu, V., Morrow, J.,
           Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
           O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
           Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
           Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
           Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
           Tesfaye, S., Theodore, J., Tirelli, A., Travers, M., Trigilio, J.,
           Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M. J.,
           Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE      Direct Submission
COMMENT     Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           On Jul 29, 2001 this sequence version replaced g1:15022095.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html

           Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu

           Project Information
           Center project name: l8056
           Center clone name: 5_O_16

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 6904: contig of 6904 bp in length
* 6905 7004: gap of 100 bp
* 7005 24083: contig of 17079 bp in length
* 24084 24183: gap of 100 bp
* 24184 74907: contig of 50724 bp in length
* 74908 75007: gap of 100 bp
* 75008 199324: contig of 124317 bp in length
* 199325 199424: gap of 100 bp
* 199425 203674: contig of 4250 bp in length.
*
* Location/Qualifiers
1. 203674

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="14"
/map="14"
/clone="RP23-5016"
/clone_1lb="RPIC1-23"
/clone_1lb_c="44067 g 52872 t 405 others"
BASE COUNT 59737 a 46593 c 44067 g 52872 t 405 others
ORIGIN

```

|                          |       |              |               |               |
|--------------------------|-------|--------------|---------------|---------------|
| Query Match              | 22.8% | Score 29.4   | DB 2          | Length 203674 |
| Best Local Similarity    | 58.6% | Pred. No. 62 |               |               |
| Matches 51, Conservative |       | 0            | Mismatches 36 | Indels 0      |
|                          |       |              |               | Gaps 0        |

Qy 41 gacaaagccccgcagccgagatgagccggagctcttttttttttggagcgtgtaacggtgttc 100  
Db 161070 GAGGATGAAAACGGGGCCAGGACGATGAGAGGTTTCTCTCCTAGACGCTGGGTGGCTTTGTA 161129

Qy 101 caacgcgtattatccgagactagtccag 127  
Db 161130 CTCGGCTCATATTAAAGGCGCTAGTGAAG 161156

| RESULT     | 10  |
|------------|---|
| AL591946   |   |
| LOCUS      |   |
| DEFINITION | AL591946 230838 bp DNA HTG 25-JUL-2001                            |
| ACCESSION  | Mus musculus chromosome 15 clone RP23-290W7, *** SEQUENCING IN    |
| VERSION    | PROGRAMS ***, in unsorted pieces.                                 |
| KEYWORDS   | AL591946  |
| SOURCE     | GI:15022302   |
| ORGANISM   | HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.                       |
|            | house mouse.  |
|            | Mus musculus  |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| AUTHORS    | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus |
| TITLE      | 1 (bases 1 to 230838)   |
| JOURNAL    | Sims, S.  |
|            | Direct Submission   |
|            | Submitted (24-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,   |
|            | CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk            |
|            | requests: clonerequests@sanger.ac.uk                              |
| COMMENT    | On Jul 26, 2001 this sequence version replaced gi:15021038.       |

```

Assembly program: XAP4, version 4.5
Sequencing vector: plasmid, L08752, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Consensus quality: 227470 bases at least Q40
Consensus quality: 228370 bases at least Q30
Consensus quality: 228935 bases at least Q20
Insert size: 229738, sum-of-contigs
Insert size: 212606, 9.8% error; agarose-fp
Quality coverage: 8.27x in Q20 bases; sum-of-contigs
Quality coverage: 9.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
  be preserved.

```

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| SOURCE   | 1. .230838          |

misc\_feature

```

misc_feature      /note="assembly_fragment:04917
                  fragment_chain:1"
                  20455. 25820
misc_feature      /note="assembly_fragment:02648
                  fragment_chain:1"
                  25921. 44695
misc_feature      /note="assembly_fragment:00262
                  fragment_chain:1"
                  44756. 78908
misc_feature      /note="assembly_fragment:05121
                  fragment_chain:1"
                  79009. 110859
misc_feature      /note="assembly_fragment:01341
                  fragment_chain:1"
                  110360. 129541
misc_feature      /note="assembly_fragment:01437
                  fragment_chain:1"
                  129642. 147588
misc_feature      /note="assembly_fragment:03096
                  fragment_chain:1"
                  147689. 168917
misc_feature      /note="assembly_fragment:02255
                  fragment_chain:1"
                  169018. 182699
misc_feature      /note="assembly_fragment:04888
                  fragment_chain:1"
                  182800. 188497
misc_feature      /note="assembly_fragment:00978
                  fragment_chain:1"
                  188598. 197147
misc_feature      /note="assembly_fragment:03681
                  fragment_chain:1"
                  197248. 230838
misc_feature      /note="assembly_fragment:03785"
                  /note="assembly_fragment:03785"
BASE COUNT      51951 a 38784 c 62500 g 56499 t 1104 others
ORIGIN

```

| Query Match | Best Local Similarity | Score | DB         | length |
|-------------|-----------------------|-------|------------|--------|
| 22.8%       | 22.8%                 | 29.4  | 2          | 230838 |
| 54          | 54                    | 0     | Indels     | 0      |
| 54          | 54                    | 0     | Mismatches | 0      |
| 54          | 54                    | 0     | Gaps       | 0      |

| RESULT     | 11  |
|------------|---|
| AF176007   |   |
| LOCUS      | AF176007 1877 bp mRNA   |
| DEFINITION | Mus musculus domesticus mitochondrial carrier homolog 1 isoform a |
| ACCESSION  | AF176007  |
| VERSION    | AF176007.2  |
| KEYWORDS   | GI:6453808  |
| SOURCE     | western European house mouse.                                     |
| ORGANISM   | Mus musculus domesticus   |

| REFERENCE<br>AUTHORS<br>TITLE<br>JOURNAL   | REFERENCE<br>AUTHORS<br>TITLE<br>JOURNAL   |
|--|--|
| 1 (bases 1 to 1877)<br>Jang, J.S., Hahn, Y., Park, C. and Chung, J.H.<br>Identification of an evolutionary conserved mitochondrial carrier<br>family from various organisms<br>Unpublished | 2 (bases 1 to 1877)<br>Jang, J.S., Hahn, Y., Park, C. and Chung, J.H.<br>Direct Submission<br>Submitted (06-AUG-1999) Department of Biological Sciences, Korea<br>Advanced Institute of Science and Technology, 373-1 Guseong-dong |

|                              |   |
|------------------------------|---|
| <b>REFERENCE</b>             | Yusong-gu, Taejon 305-701, South Korea  |
| <b>AUTHORS</b>               | 3 (bases 1 to 1877)   |
| <b>TITLE</b>                 | Jang,J.S., Hahn,Y. and Chung,J.H.   |
| <b>JOURNAL</b>               | Direct Submission   |
| <b>REMARK</b>                | Submitted (19-Nov-1999) Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong Yusong-gu, Taejon 305-701, South Korea  |
| <b>COMMENT</b>               | Sequence update by submitter  |
| <b>FEATURES</b>              | On Nov 19, 1999 this sequence version replaced gi:5815342.  |
| <b>SOURCE</b>                | Location/Qualifiers<br>1..1877<br>"/organism='Mus musculus domesticus'<br>'/db_species='domesticus'<br>'/gb_xref='taxon:10092'<br>119..1237<br>'/note='MtcHla'<br>'/codon_start=1<br>'/product='mitochondrial carrier homolog 1 isoform a'<br>'/protein_id='AAD52645.2'<br>'/db_xref='gi:5453809'<br>'/translation='MGASDPVAPWAPGAGAGACAGARCGAPAGVEARADPP<br>PANHARHPRAPOPSARRMDGGGAGSGSDNPTPEALPAVGATLASHPLLV<br>KLILVOGHPEMPPLTGVNLGRKLYLLSPFTYAKTIYQNGKIGIRRGISPRLSNNA<br>LSYTRGSMKRVPRPDEMEQVSNNDDKTSIAKTIVKEYSIMMQCSRLHAFLHYIT<br>SMRQVDVPGREAKITSGLVSLTGIFIEBGLGFVGLIPLILDVVFLWCNLIARE<br>INAILVDSEFSQLAIRSYTKFVMGIASVMLTYPLFVLGDLMAVNNGLRAGLPDYS<br>VFESMKHWKXLSVQGLFRGSILLFRVSSGSCFALE'   |
| <b>POLYA_SIGNAL</b>          | polya_signal<br>polya_site     1867     563 c     578 g     404 t   |
| <b>BASE COUNT</b>            | 332 a     563 c     578 g     404 t   |
| <b>ORIGIN</b>                |   |
| <b>Query Match</b>           | 22.6%; Score 29.2; DB 10; Length 1877;  |
| <b>Best Local Similarity</b> | 59.8%; Pred. No. 62;  |
| <b>Matches</b>               | 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  |
| <b>OY</b>                    | 48 accccgaccgagatgctgcgggctcctttgttgtgtgctgtgacgtgtgtccaacct 107<br>  |
| <b>Db</b>                    | 887 ATCCCTCACCTTCGTGGGAGATGTTGTTCTGTGGGCTGTAAACTGCTGCCCACTTC 946<br>  |
| <b>OY</b>                    | 108 attattccgactagtcaagg 129<br>  |
| <b>Db</b>                    | 947 ATCATGCCTACTTGCTGGACC 968<br>   |
| <b>RESULT 12</b>             |   |
| <b>AF192558</b>              | 1928 bp mRNA ROD 26-NOV-1999  |
| <b>LOCUS</b>                 | Mus musculus domesticus mitochondrial carrier homolog 1 isoform b   |
| <b>DEFINITION</b>            | (MTC1) mRNA, complete cds; nuclear gene for mitochondrial product.  |
| <b>ACCESSION</b>             | AF192558  |
| <b>VERSION</b>               | AF192558.2 GI:6468759   |
| <b>KEYWORDS</b>              |   |
| <b>SOURCE</b>                | western European house mouse.   |
| <b>ORGANISM</b>              | Mus musculus domesticus   |
| <b>REFERENCE</b>             | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;<br>Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.<br>1 (bases 1 to 1928)<br>Jang,J.S., Hahn,Y., Park,C. and Chung,J.H.<br>Evolutionarily conserved mitochondrial carrier gene family in<br>mouse, human, and zebrafish<br>Unpublished<br>2 (bases 1 to 1928)<br>Jang,J.S., Hahn,Y., Park,C. and Chung,J.H.<br>Direct Submission<br>Submitted (06-Oct-1999) Department of Biological Sciences, Korea<br>Advanced Institute of Science and Technology, 373-1 Guseong-dong<br>Yusong-gu, Taejon 305-701, South Korea<br>3 (bases 1 to 1928)<br>Lee,J., Hahn,Y. and Chung,J.H.<br>Direct Submission<br>Submitted (26-Nov-1999) Department of Biological Sciences, Korea |
| <b>REFERENCE</b>             |   |
| <b>AUTHORS</b>               |   |
| <b>TITLE</b>                 |   |
| <b>JOURNAL</b>               |   |

| REMARK  | COMMENT  | FEATURES   | source |
|---|--|--|--------|
| Advanced Institute of Science and Technology, 373-1 Gusong-dong<br>Yuseong-gu, Daejeon 305-701, South Korea | Sequence update by submitter<br>On Nov 26, 1999 this sequence version replaced gi:6466115. | Location/Qualifiers  |        |
|   |  | 1..1928  |        |
|   |  | /organism="Mus musculus domesticus"  |        |
|   |  | /sub_species="domesticus"  |        |
|   |  | /db_xref="taxon:10092"   |        |
|   |  | 1..1928  |        |
| gene  |  | /gene="Mtchl"  |        |
|   |  | 119..1288  |        |
| CDS   |  | /gene="Mtchl"  |        |
|   |  | /codon_start=1   |        |
|   |  | /product="mitochondrial carrier homolog 1 isoform b"   |        |
|   |  | /protein_id="AAFI2792.2"   |        |
|   |  | /db_xref="GI:6468760"  |        |
|   |  | /translation="MGASDEPEVAPWAPGAGMAGAGACAGARGAPVGEARARDDPP<br>PAHRAHGHPRHRAAPQSRARMDGGVAGAGSGDNAPTEPALEPAVAGATGATLSHPLTV<br>KLIOVGHPEHPHPIGTNVIGIKVLYIDPPTVAKYIVQDNGKIGISRGILSRPLMSNN<br>LSVTYRSGMKKVFPPDDEMDVSNKDDKTSLSKTKVKEISYEMMOCSIRBLAHPLHTY<br>SMKRVQFVGEAREATISGVLSIGKIFEEBGLGFEVGLIPHLGSDVVFWMCCNLIAHFI<br>INATLVDSVSDTEKPGGJGNDONPGSOFQALAIISYKFEVGVIAVSMLTYPELLVGDI<br>MAVNVNGICRAELPSPYSPVKSMIHCWKYLSVQGOLFRRSSILFRVSSGSCFALE" |        |
|   |  | 1289..1925   |        |
| 3' UTR  |  | /gene="Mtchl"  |        |
|   |  | 1894..1899   |        |
| polyA_signal  |  | /gene="Mtchl"  |        |
|   |  | 1918   |        |
| polyA_site  |  | /gene="Mtchl"  |        |
|   |  | 345 a 577 c 596 g 410 t  |        |
| BASE COUNT  |  |  |        |
| ORIGIN  |  |  |        |

| RESULT     | 13  |
|------------|---|
| LOCUS      | RLU44387  |
| DEFINITION | Rhizobium leguminosarum bv. trifolii ORF1 gene, complete cds, ORF2 gene, partial cds.   |
| ACCESSION  | U44387  |
| VERSION    | U44387.1  |
| KEYWORDS   | GI:1172153  |
| SOURCE     |   |
| ORGANISM   | Rhizobium leguminosarum bv. trifolii.<br>Rhizobium leguminosarum bv. trifolii.<br>Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.   |
| REFERENCE  | 1 (bases 1 to 2193)<br>Krol, J., and Skorupska, A.<br>Identification of genes in Rhizobium leguminosarum bv. trifolii whose products are homologues to a family of ATP-binding proteins Microbiology 143 (Pt 4), 1389-1394 (1997)   |
| JOURNAL    | Medline   |
| REFERENCE  | 2 (bases 1 to 2193)<br>Krol, J.E. and Skorupska, A.M.<br>Direct Submission<br>Submitted (29-DEC-1995) Jaroslaw E. Krol, General Microbiology, M Curie-Sklodowska University, Akademicka 19, Lublin, 20-033, Poland<br>Cloned on cosmid pARF136 complementing Exo-mutations, Skorupska, A., Acta Biochim. Pol. 38, 423-435 (1991). |
| COMMENT    |   |



FEATURES  
SOURCE

Location/Qualifiers  
1.2193  
/organism="Rhizobium leguminosarum bv. trifolii"  
/strain="TA1"  
/db\_xref="taxon:386"  
41.44  
46.1758  
/note="ORF1, similar to product encoded by Pseudomonas aeruginosa aprd gene, GenBank Accession Number X64558 and to product encoded by Erwinia chrysanthemi prtd gene, GenBank Accession Number M60395"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAB57756.1"  
/db\_xref="GI:1172154"  
/translation="MOGLVIFGIALINILYTSFPMLEVDRIPIPSISLAY LAIALMLYCPGAFELRSRLVAVGALDEWNGRFRALIKAPLVKIGDGLP LRDPIQRTPLSGMPTAFDLPMLPYIVICFLPHRAIGTAIGSLVALITMTN OGRTLSKQSDSANMNAFAOTSIRNSEVHAMAGTMAEIMDRKSEYRTITROA SDVNGVATLSKIFRIALQSGTLATGALIVGQSSGIIAGSILTSRALAPVAAI GNMRFVSAOOSMARLSLKTIPETIPAPLAAPTQVWEGALGCPAGORLIVSD VSFGRAGSALGVIGYSASGKSLARAMGIMPTVAGSIRLDGALDMDGALGRHI GYLPODVELEFGTYAONICRFAKESPEVYVAAARAVHDLILPLNGEYETEGG AALSAGORRIALALYGEPLIVLDPEPNSLDEGERALSAIMSVRAGGIYVI AHRSGVLAFCDFVLMQEGRMIAFGPKBEVLARVSRPEARTPIAERVAOLKVVDDM NAAE"  
1764.1770  
/note="ORF2, similar to product encoded by Pseudomonas aeruginosa aprd gene, GenBank Accession Number X64558 and to product encoded by Erwinia chrysanthemi prtd gene, GenBank Accession Number M60395"  
/codon\_start=1  
/transl\_table=11  
/product="unknown"  
/protein\_id="AAB57757.1"  
/db\_xref="GI:1172155"  
/translation="MSKVISSEKSLNRHNAVGVLSIALVCGIGGMAATLSSAVI GEGVIVDPKCONIHLTGIVSELVSENDHVTAGVILRLDGTTRANLSIYESTL AQLYARRARLAEIRIGAESFEVEENTIDLTSTSAOKL"  
596 c 684 g 485 t

BASE COUNT 428 a 596 c 684 g 485 t

ORIGIN

Query Match 22.6%; Score 29.2; DB 1; Length 2193;  
Best Local Similarity 57.8%; Pred. No. 62;  
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 28 gagcattctgacgaacacccgacgagatgctggggctctttgtgtgtgct 87  
DB 374 GAACCTTCTCGCGGTATGGCCGACGACGATCTGCATCTGCCGCTGCGTCT 433  
QY 88 gtacagtgtgtccacacgtattatccgg 117  
DB 434 ATATGCTATCTCTCTCTCTTCAATCCG 463

RESULT 14

LOCUS 5263 bp DNA BCT 29-OCT-1997  
DEFINITION Rhizobium leguminosarum prsd, prse, ORF3 genes.  
VERSION X98117.1 GI:1806273  
KEYWORDS ORF3 gene; prsd gene; prse gene.  
SOURCE Rhizobium leguminosarum.  
ORGANISM Rhizobium leguminosarum.  
Bacteria: Proteobacteria; alpha subdivision: Rhizobiaceae group; Rhizobiaceae; Rhizobium.  
Krol, J. and Skoriupska, A.  
REFERENCE 1 (bases 1 to 5263)  
AUTHORS Krol, J. and Skoriupska, A.  
TITLE Identification of genes in Rhizobium leguminosarum bv. trifolii whose products are homologues to a family of ATP-binding proteins

JOURNAL Microbiology 143 (Pt 4), 1389-1394 (1997)  
MEDLINE 97286554  
REFERENCE 2 (bases 1 to 4423)  
AUTHORS Krol, J. E.  
TITLE Direct Submission  
COMMENT Submitted (24-May-1996) J.E. Krol, Maria Curie Sklodowska University - Lublin, Dept. of General Microbiology, Akademicka 19, PL-20 033 Lublin, POLAND  
On Jan 29, 1997 this sequence version replaced gi:1360123.  
FEATURES  
SOURCE

1.5263  
/organism="Rhizobium leguminosarum"  
/sub\_species="trifolii"  
/strain="TA1"  
/db\_xref="taxon:384"  
/clone\_1lb="PARF136"  
/clone="PARF136"  
/sub\_clone="pbf1"  
1250.1254  
1250.1254  
1256.2968  
/gene="prsd"  
1256.2968  
/gene="prsd"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA66799.1"  
/db\_xref="GI:1360124"  
/db\_xref="SPTREMBL:057040"  
/translation="MOGLVIFGIALINILYTSFPMLEVDRIPIPSISLAY LAIALMLYCPGAFELRSRLVAVGALDEWNGRFRALIKAPLVKIGDGLP LRDPIQRTPLSGMPTAFDLPMLPYIVICFLPHRAIGTAIGSLVALITMTN OGRTLSKQSDSANMNAFAOTSIRNSEVHAMAGTMAEIMDRKSEYRTITROA SDVNGVATLSKIFRIALQSGTLATGALIVGQSSGIIAGSILTSRALAPVAAI GNMRFVSAOOSMARLSLKTIPETIPAPLAAPTQVWEGALGCPAGORLIVSD VSFGRAGSALGVIGYSASGKSLARAMGIMPTVAGSIRLDGALDMDGALGRHI GYLPODVELEFGTYAONICRFAKESPEVYVAAARAVHDLILPLNGEYETEGG AALSAGORRIALALYGEPLIVLDPEPNSLDEGERALSAIMSVRAGGIYVI AHRSGVLAFCDFVLMQEGRMIAFGPKBEVLARVSRPEARTPIAERVAOLKVVDDM NAAE"  
2975.2981  
2984.4285  
2984.4285  
/gene="prse"  
/gene="prse"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CA66800.1"  
/db\_xref="GI:1360125"  
/db\_xref="SPTREMBL:052859"  
/translation="MSKVISSEKSLNRHNAVGVLSIALVCGIGGMAATLSSAVI GEGVIVDPKCONIHLTGIVSELVSENDHVTAGVILRLDGTTRANLSIYESTL AQLYARRARLAEIRIGAESFEVEENTIDLTSTSAOKLDEOKLFLPFRSALIGMS OLASRKQALAEQIKGLVVOINATNDSGLIEEELGSDIVLYKKGLVLPFSAGAC RSPOORPBGNSKSGGEGKALIEDROSQLEDRESEAKUTLVEAOIAFYERAEF RSICIVTSRAADRTSSRAVHTVNGVIDGUTLMLVPENNELFYEAKVATDIDQ VHVQSDVRFSAFDQRTTPDVSGEITISADIYKDEKRTGISYPLRVKPKPESIAK KTIKLPGMPAEVFIKIGDRVTSYILKPLTDMQHVFQF"

BASE COUNT 1123 a 1430 c 1553 g 1157 t

ORIGIN



|               |  |
|---------------|--|
| intron        | /standard_name="GC_rich<br>/note="low_complexity"<br>5663..10252<br>/gene="Mchl"   |
| repeat_region | /number=1<br>6800..6946<br>/standard_name="B1_MW"<br>/note="SINE"<br>/rpt_family="Alu"<br>8270..8334<br>/standard_name="B1_MW"<br>/note="SINE"<br>/rpt_family="Alu"<br>9109..9257<br>/standard_name="SINE1"<br>/note="SINE"<br>/rpt_family="B2"<br>9451..9487<br>/rpt_type=" tandem<br>/rpt_unit="caaa<br>9488..9510<br>/standard_name="B2"<br>/note="SINE"<br>/rpt_family="B2"<br>9516..9596<br>/standard_name="PB1D9"<br>/note="SINE"<br>/rpt_family="Alu"<br>9597..9625<br>/rpt_type=" tandem<br>/rpt_unit="caaaa<br>9633..9804<br>/standard_name="SINE1"<br>/note="SINE"<br>/rpt_family="B4"<br>10253..10337<br>/gene="Mchl" |
| intron        | /number=2<br>10338..12257<br>/gene="Mchl"  |
| exon          | /number=2<br>12258..12364<br>/gene="Mchl"<br>/number=3<br>12365..12657<br>/gene="Mchl"<br>/number=3<br>12658..12735<br>/gene="Mchl"<br>/number=4<br>12736..13027<br>/gene="Mchl"   |
| intron        | /number=4<br>13028..13085<br>/gene="Mchl"  |
| exon          | /number=5<br>13086..13310<br>/gene="Mchl"  |
| intron        | /number=5<br>13311..13362<br>/gene="Mchl"  |
| exon          | /number=6<br>13363..14350<br>/gene="Mchl"  |
| intron        | /number=6<br>14351..14410  |

|                       |        |               |        |               |
|-----------------------|--------|---------------|--------|---------------|
| Query Match           | 22.68; | Score 29.2;   | DB 10; | Length 21984; |
| Best Local Similarity | 59.88; | Pred. NO. 67; |        |               |

|    | Matches | 49;  | Conservative | 0; | Mismatches | 33; | Indels | 0; | Gaps | 0; |
|----|---------|--|--------------|----|------------|-----|--------|----|------|----|
| QY | 48      | accaccgacacgagatgctcgaggctctttgtgtgtgctgtagcgtgtgtccaacgt    | 107          |    |            |     |        |    |      |    |
| Db | 16844   | ATCCCTCACACCCCTGGGGGAGATGTGTTTCTTGTGTGGGCTGTAACTCTGGCCACATTC | 16903        |    |            |     |        |    |      |    |
| QY | 108     | attatttcggaactgaattccagcg                                    | 129          |    |            |     |        |    |      |    |
| Db | 16904   | ATCATAGCCTACTTGTGTGAGCG                                      | 16925        |    |            |     |        |    |      |    |

Search completed: March 28, 2002, 15:20:21  
Job time: 8185 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:15:31 ; Search time 240.01 Seconds

(Without alignments)  
460.793 Million cell updates/sec

Title: US-09-816-391a-1\_COPY\_472\_600

Perfect score: 129

Sequence: 1 tgaacttctgcgcgacgca.....tattcgcgactagtcagcg 129

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.1101.\*

|     |                                      |
|-----|--------------------------------------|
| 1:  | /SIDS2/gcgcdata/geneseq/NA1980.DAT.* |
| 2:  | /SIDS2/gcgcdata/geneseq/NA1981.DAT.* |
| 3:  | /SIDS2/gcgcdata/geneseq/NA1982.DAT.* |
| 4:  | /SIDS2/gcgcdata/geneseq/NA1983.DAT.* |
| 5:  | /SIDS2/gcgcdata/geneseq/NA1984.DAT.* |
| 6:  | /SIDS2/gcgcdata/geneseq/NA1985.DAT.* |
| 7:  | /SIDS2/gcgcdata/geneseq/NA1986.DAT.* |
| 8:  | /SIDS2/gcgcdata/geneseq/NA1987.DAT.* |
| 9:  | /SIDS2/gcgcdata/geneseq/NA1988.DAT.* |
| 10: | /SIDS2/gcgcdata/geneseq/NA1989.DAT.* |
| 11: | /SIDS2/gcgcdata/geneseq/NA1990.DAT.* |
| 12: | /SIDS2/gcgcdata/geneseq/NA1991.DAT.* |
| 13: | /SIDS2/gcgcdata/geneseq/NA1992.DAT.* |
| 14: | /SIDS2/gcgcdata/geneseq/NA1993.DAT.* |
| 15: | /SIDS2/gcgcdata/geneseq/NA1994.DAT.* |
| 16: | /SIDS2/gcgcdata/geneseq/NA1995.DAT.* |
| 17: | /SIDS2/gcgcdata/geneseq/NA1996.DAT.* |
| 18: | /SIDS2/gcgcdata/geneseq/NA1997.DAT.* |
| 19: | /SIDS2/gcgcdata/geneseq/NA1998.DAT.* |
| 20: | /SIDS2/gcgcdata/geneseq/NA1999.DAT.* |
| 21: | /SIDS2/gcgcdata/geneseq/NA2000.DAT.* |
| 22: | /SIDS2/gcgcdata/geneseq/NA2001.DAT.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match Length | ID        | Description                 |
|------------|-------|--------------------|-----------|-----------------------------|
| 1          | 28.4  | 22.0               | 936 18    | AA79329 DNA encoding Arch   |
| 2          | 27.6  | 21.4               | 366 18    | AA767819 H. pylori cell env |
| 3          | 27.6  | 21.4               | 366 18    | AA77499 H. pylori inner me  |
| 4          | 27.6  | 21.4               | 1422 20   | AA35720 CDNA encoding a pr  |
| 5          | 27.6  | 21.4               | 1613 20   | AA35721 CDNA encoding a pr  |
| 6          | 27.6  | 21.4               | 2090 19   | AAV68588 Nucleotide sequenc |
| 7          | 27.6  | 21.4               | 2090 19   | AAV38564 Human nuclear prot |
| 8          | 27.6  | 21.4               | 2150 22   | AAH81781 Human differential |
| 9          | 27.6  | 21.4               | 26778 21  | AAH81477 N. meningitidis pa |
| 10         | 27.6  | 21.4               | 349980 21 | AAE21612 Neisseria meningit |
| 11         | 27.4  | 21.2               | 1356 22   | AA521265 Human CDNA sequenc |

|    |      |      |         |          |
|----|------|------|---------|----------|
| 12 | 27.4 | 21.2 | 1670 21 | AA233333 |
| 13 | 27.4 | 21.2 | 1878 22 | AAH16020 |
| 14 | 27.4 | 21.2 | 1951 20 | AA337414 |
| 15 | 27.2 | 21.1 | 2628 22 | AAH01309 |
| 16 | 26.8 | 20.8 | 330 21  | AA444633 |
| 17 | 26.8 | 20.8 | 364 21  | AA38294  |
| 18 | 26.8 | 20.8 | 418 21  | AA46662  |
| 19 | 26.4 | 20.5 | 927 21  | AA43424  |
| 20 | 26.4 | 20.5 | 1364 21 | AA45331  |
| 21 | 26.4 | 20.5 | 1365 21 | AA39325  |
| 22 | 26.2 | 20.3 | 446 22  | AA112040 |
| 23 | 26.2 | 20.3 | 446 22  | AA13374  |
| 24 | 26.2 | 20.3 | 446 22  | AA101966 |
| 25 | 26.2 | 20.3 | 598 22  | AA10947  |
| 26 | 26.2 | 20.3 | 746 21  | AA92283  |
| 27 | 26.2 | 20.3 | 907 19  | AAV20670 |
| 28 | 26.2 | 20.3 | 907 19  | AAV20670 |
| 29 | 26.2 | 20.3 | 927 20  | AAV20669 |
| 30 | 26.2 | 20.3 | 1001 19 | AAV20669 |
| 31 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 32 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 33 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 34 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 35 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 36 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 37 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 38 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 39 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 40 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 41 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 42 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 43 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 44 | 26.2 | 20.2 | 1173 22 | AAV20669 |
| 45 | 26.2 | 20.2 | 1173 22 | AAV20669 |

## ALIGNMENTS

|          |  |                         |
|----------|--|-------------------------|
| RESULT 1 | AA79329  | standard; DNA; 936 BP.  |
| ID       | AA79329  | standard; DNA; 936 BP.  |
| AC       | AA79329  |                         |
| XX       | AA79329  |                         |
| XX       | AA79329  |                         |
| DT       | 16-FEB-1998  | (first entry)           |
| DE       | DNA encoding Archaeoglobus fulgidus esterase VC16-16NC.        |                         |
| XX       | DNA encoding Archaeoglobus fulgidus esterase VC16-16NC.        |                         |
| XX       | DNA encoding Archaeoglobus fulgidus esterase VC16-16NC.        |                         |
| KW       | Esterase; thermostable enzyme; ester; chiral compound; cheese; |                         |
| KW       | pulp; paper; lignin removal; sugar; lignocellulose;            |                         |
| KW       | disease resistance; feedstuff; ss.                             |                         |
| XX       | Archaeoglobus fulgidus strain VC16.                            |                         |
| OS       | Archaeoglobus fulgidus strain VC16.                            |                         |
| XX       | Archaeoglobus fulgidus strain VC16.                            |                         |
| XX       | Archaeoglobus fulgidus strain VC16.                            |                         |
| FT       | Key  | Location/Qualifiers     |
| FT       | unsure   | 313..315                |
| FT       | CDS  | 1..936                  |
| FT       | /*tag- a   | /note="encodes Ile"     |
| FT       | /*tag- b   |                         |
| FT       | /transl_except-  | (pos: 382..384, aa:Ala) |
| FT       | /transl_except-  | (pos: 385..387, aa:Ala) |
| FT       | /transl_except-  | (pos: 430..432, aa:Ala) |
| FT       | /transl_except-  | (pos: 487..489, aa:Gly) |
| FT       | /transl_except-  | (pos: 499..501, aa:Ala) |
| FT       | /transl_except-  | (pos: 505..507, aa:Val) |
| FT       | /transl_except-  | (pos: 559..561, aa:Ile) |
| FT       | /transl_except-  | (pos: 691..693, aa:Asn) |
| PN       | W09730160-A1.  |                         |
| XX       | 21-AUG-1997.   |                         |
| PD       | 21-AUG-1997.   |                         |

Human secreted pro  
Human CDNA sequenc  
Human secreted pro  
Escherichia coli n  
Arabidopsis thalia  
Zea mays DNA fragm  
Arabidopsis thalia  
Arabidopsis thalia  
Probe #1973 for ge  
Probe #2060 used t  
Probe #9633 used t  
Porcine melanocort  
Human ninturin 2 v  
Human ninturin 2 e  
Human secreted pro  
Human ninturin 2 v  
Pseudomonas stutze  
H. pylori inner me  
H. pylori outer me  
Human bone marrow  
Pseudomonas stutze  
Tyrlactone synthase  
Beta-carotene hydr  
Alcaligenes (Deley  
Mouse SM1 gene. M  
Human Down syndrom  
Zea mays DNA fragm  
Human MCP-4 recept  
Human CCR-2 CC che  
Human adenosine re

XX 11-FEB-1997; 97MO-US02039.  
 XX 16-FEB-1996; 96US-0602359.  
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 XX Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;  
 PI Reid J, Robertson DE, Swanson RV, Warren PV;  
 DR WPI: 1997-425035/39.  
 DR P-PSDB; AAM23076.  
 PT Nucleic acid encoding heat stable esterase from thermophilic  
 PT bacteria - which is active in organic solvents, useful in cheese or  
 PT paper manufacture, and to study plant resistance to disease  
 XX  
 XX Claim 1; Page 50-51; 113pp; English.  
 CC This DNA sequence codes for thermostable esterase VCI6-16MC  
 CC (AAM23076) of *Archaeoglobus fulgidus* VCI6, an isolate that grows  
 CC optimally at 85 deg C and pH 7.0. It can be amplified from a  
 CC plasmid vector by PCR (see AAT79317-18). Claimed, newly  
 CC identified polynucleotides (AAT79321-30) encoding esterases  
 CC (AAM23069-77, AAM23088) were recovered from genomic gene libraries.  
 CC They can be used for recombinant production of the enzymes in host  
 CC cells, and as probes to identify related sequences. The esterases  
 CC are stable at high temperature and in organic solvents, making them  
 CC superior for use in production of pure chiral compounds used in  
 CC pharmaceutical, agricultural and other chemical industries. A  
 CC method is claimed for transferring an amino group from an amino  
 CC acid to an alpha-keto acid using a claimed esterase. The enzymes  
 CC may also be useful as ripening starters in cheese making, in lignin  
 CC removal in paper and pulp manufacture, in carbohydrate derivative  
 CC synthesis, in fermentable sugar production from lignocellulosic  
 CC waste, in the study of plant wall structure, plant resistance to  
 CC disease and organic matter decomposition and to select plants bred  
 CC for production of highly degradable animal feeds.  
 XX  
 XX Sequence 936 BP; 235 A; 220 C; 265 G; 214 T; 2 other;  
 S0  
 Query Match 22.0%; Score 28.4; DB 18; Length 936;  
 Best Local Similarity 56.4%; Pred. No. 2.9;  
 Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;  
 QY 2 gactctctgctagcagattactgagcattactgacgaacacccgcagagat 61  
 Db 718 gactctgagacattactcctcctgctgctgatacaccgcgaatacaccgctgagagat 777  
 QY 62 ggtcggggtcttttctgtgtgtgctgtgagctg 95  
 Db 778 gaagagaagattctcggcagatgctgagaagag 811  
 RESULT 2  
 AAT67819  
 ID AAT67819 standard; DNA; 366 BP.  
 XX  
 XX AAT67819;  
 XX 04-AUG-1997 (first entry)  
 DE H. pylori cell envelope inner membrane protein ORF 486075.aa.  
 XX  
 XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis; ds.  
 XX  
 XX Helicobacter pylori.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH

FT CDS 1..366  
 FT /tag- a  
 FT /transl\_except- pos: 289..291, aa: Xaa  
 FT /transl\_except- pos: 301..303, aa: Xaa  
 FT /transl\_except- pos: 304..306, aa: Xaa  
 FT /transl\_except- pos: 316..318, aa: Xaa  
 FT /transl\_except- pos: 346..348, aa: Xaa  
 FT /transl\_except- pos: 355..357, aa: Xaa  
 FT /transl\_except- pos: 361..363, aa: Xaa  
 FT /note- "Xaa-unknown, no start codon is given"  
 XX  
 XX MO9640893-A1.  
 XX  
 XX 19-DEC-1996.  
 XX  
 XX 06-JUN-1996; 96MO-US09122.  
 XX  
 XX 01-APR-1996; 96US-0630405.  
 XX 07-JUN-1995; 95US-0487032.  
 XX  
 XX (ASTR ) ASTRA AB.  
 XX  
 XX Berglindh OT, Smith D, Mellgaerd BL;  
 XX  
 XX WPI: 1997-052306/05.  
 XX P-PSDB; AAM20524.  
 DR  
 XX  
 XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 PT  
 PS Claim 1; Page 7; 1481pp; English.  
 XX  
 XX This sequence encodes a Helicobacter pylori cell envelope protein  
 CC that may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 CC Note: This DNA sequence is not reproduced in the specification and  
 CC has been derived from the related specification, WO9719098.  
 CC  
 XX  
 S0 Sequence 366 BP; 109 A; 62 C; 79 G; 109 T; 7 other;  
 Query Match 21.4%; Score 27.6; DB 18; Length 366;  
 Best Local Similarity 58.5%; Pred. No. 4;  
 Matches 48; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
 QY 19 gattactctgagcattactgacgaacacccgcagagatgctggtcttttctgt 78  
 Db 93 gactccttgatcattgaaagaacagatgacgaatggaatgcatgctgttctgt 152  
 QY 79 tgtggtgctgagcgtgtgtc 100  
 Db 153 ggtggtgtgaaagcagtgctatc 174  
 RESULT 3  
 AAT77499  
 ID AAT77499 standard; DNA; 366 BP.  
 XX  
 XX AAT77499;  
 XX 11-AUG-1997 (first entry)  
 XX

XX H. pylori inner membrane protein ORF 486075.aa.  
XX  
XX Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
KW detection; antisense; inhibition; ds.  
XX  
OS Helicobacter pylori.  
XX  
FH Key location/Qualifiers  
FH CDS 1..366  
FT /\*tag-  
FT a  
FT /transl\_except- (pos:289..291, aa:Xaa)  
FT /transl\_except- (pos:301..303, aa:Xaa)  
FT /transl\_except- (pos:304..306, aa:Xaa)  
FT /transl\_except- (pos:316..318, aa:Xaa)  
FT /transl\_except- (pos:346..348, aa:Xaa)  
FT /transl\_except- (pos:355..357, aa:Xaa)  
FT /transl\_except- (pos:361..363, aa:Xaa)  
XX  
XX MO9719098-A1.  
XX  
XX 29-MAY-1997.  
XX  
XX 15-NOV-1996; 96MO-US18542.  
XX  
XX 17-NOV-1995; 950S-0561469.  
XX  
XX (ASTR ) ASTRA AB.  
XX  
XX Smith DH;  
XX  
XX MPI: 1997-298052/27.  
XX  
XX P-PSDB; AAM24680.  
XX  
XX Helicobacter pylori nucleic acid sequences and related proteins -  
XX used for diagnostics and therapeutics  
XX  
XX Claim 1; Page 122; 235pp; English.  
XX  
XX This sequence encodes a Helicobacter pylori inner membrane protein.  
XX Helicobacter pylori has been strongly linked to chronic gastritis and  
XX duodenal ulcer disease. The nucleic acid sequences of the invention  
XX are used to evaluate compounds, especially activators or inhibitors of  
XX bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
XX sequence. The nucleic acid sequences, and corresponding proteins, are  
XX also useful for generating vaccines for immunising subjects against H.  
XX pylori or for use in detecting the presence of Helicobacter species in  
XX a sample. Antisense nucleic acid sequences of these sequences are  
XX used to inhibit expression of a gene from Helicobacter species. H.  
XX pylori whole genomic DNA was isolated and nebulised to a median size of  
XX 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
XX BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
XX complementary to the BstXI-cut pmx vectors, while the overhang is not  
XX self-complementary. Therefore the linkers will not concatemise nor  
XX will the cut vector re-ligate itself easily. The linker-adaptor inserts  
XX were ligated to each of the 20 pmx vectors to construct a series of  
XX shotgun subclone libraries. The purified DNA samples were then  
XX sequenced.  
XX Note: The ORF/protein reference number for this sequence was obtained  
XX from the related specification, WO9640893.  
XX  
XX Sequence 366 BP; 109 A; 62 C; 79 G; 109 T; 7 other;

OY 79 tttgtgctgttgacgttctgc 100  
Db 153 gtttgtcagtcagatgtalc 174

RESULT 4  
AAK35720/c  
ID AAK35720 standard; cDNA; 1422 BP.  
XX  
AC AAK35720;  
XX  
DT 09-JUL-1999 (first entry)  
DE cdna encoding a protein identified by the signal sequence trap method.

XX  
KW Signal sequence trap method; SST method; immunisation; inhibition;  
KW infection; allergy; cancer; regulation; tissue formation; tissue repair;  
KW activin activity; inhibin activity; chemokine activity;  
KW cytokine activity; blood coagulation regulation; agonist; antagonist;  
KW metabolic disorder; hormonal disorder; immune disorder;  
KW severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;  
WW wound; ss.  
OS Homo sapiens.  
PN WO9918126-A1.  
PD 15-APR-1999.  
PF 06-OCT-1998; 98WO-JP04514.  
PR 07-OCT-1997; 97JP-0274674.  
PA (ONOY ) ONO PHARM CO LTD.  
PI Fukushima D, Shibayama S, Tada H;  
DR WPI; 1999-277254/23.  
PT P-PSTB; AAY02371.  
XX Polypeptides identified by the signal sequence trap method from a  
PT human CDNA library

XX  
PS Claim 5; Page 139-140; 281pp; Japanese.  
XX  
XX AAK35694-X35747 represent cDNA sequences that encode novel polypeptides  
CC (AAY02358-84) which are identified from a human placental cDNA library  
CC by the signal sequence trap (SST) method. The polypeptides have a  
CC broad range of physiological activity, including immunisation against  
CC and inhibition of infections, allergies and cancer; regulation of tissue  
CC formation and repair; activin/inhibin activity; chemokine/cytokine  
CC activity; blood coagulation regulation; and receptor/ligand agonist  
CC or antagonist activity. The polypeptides can be used for prevention  
CC and treatment of disorders including infections by bacteria, yeasts and  
CC viruses (including HIV) and protozoa; metabolic and hormonal disorders;  
CC immune disorders (including severe combined immunodeficiency (SCID)  
CC and AIDS; thrombosis; cancer; and traumatic or surgical wounds.  
XX

Sequence 1422 BP; 538 A; 213 C; 332 G; 339 T; 0 other;

| Query Match           | 21.4% | Score 27.6  | DB 18  | Length 366    |
|-----------------------|-------|---|--------|---------------|
| Best Local Similarity | 56.5% | Pred. No. 4   |        |               |
| Matches               | 48    | Conservative  | 0      | Mismatches 34 |
|                       |       |   | Indels | 0             |
|                       |       |   |        | Gaps 0        |
| OY                    | 19    | gattactcgcgactactgacacaaagaccccgacgagttgtcgaggtctttt    | 78     |               |
|                       |       |   |        |               |
| Db                    | 93    | gatcgcttgcattcatgtaaaagacaaagctctatgacgaagtcgagtcgctttt | 152    |               |

|         | Query Match | Best Local Similarity                                       | 21.4% | Score 27.6 | DB 20 | Length 1422 |
|---------|-------------|---|-------|------------|-------|-------------|
| Matches | 60          | Conservative  | 0     | Mismatches | 54    | Indels      |
|         |             |   |       |            |       | Gaps        |
| Qy      | 7           | tctgctgtagcgaattactcagagcattacgagacagaaccccgccgagatgctcg    | 66    |            |       |             |
|         |             |   |       |            |       |             |
| Db      | 631         | TCCTCTGCATTCATTCTTTTCATGATTAATTTCGCTCAATCTCTCAACTGAGTACTTCT | 572   |            |       |             |
| Qy      | 67          | ggagctctttctgtcgtgctgcgtgaacggttgcacacgctatatttcgagct       | 120   |            |       |             |
|         |             |   |       |            |       |             |
| Db      | 571         | GATTGTCAATGATATTCTGTAGTGCAGTGTCTCCCTGTTTATATGTTTCAATT       | 518   |            |       |             |

## RESULT 5

AAK35721/c

ID AAK35721 standard; CDNA: 1613 BP.

XX AAK35721;

AC 09-JUL-1999 (first entry)

XX cDNA encoding a protein identified by the signal sequence trap method.

XX Signal sequence trap method; SST method; immunisation; inhibition;

XX infection; allergy; cancer; regulation; tissue formation; tissue repair;

XX activin activity; inhibin activity; chemokine activity;

XX cyclokin activity; blood coagulation regulation; agonist; antagonist;

XX metabolic disorder; hormonal disorder; immune disorder;

XX severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;

XX wound; ss.

XX Homo sapiens.

XX MO9918126-A1.

XX 15-APR-1999.

XX 06-OCT-1998; 98WO-JP04514.

XX 07-OCT-1997; 97JP-0274674.

XX (ONKO) ONO PHARM CO LTD.

XX Fukushima D, Shibayama S, Tada H;

XX WPI: 1999-277254/23.

XX P-PSDB: AAY02371.

XX Polypeptides identified by the signal sequence trap method from a

XX human cDNA library

XX Claim 4; Page 141-144; 281pp; Japanese.

XX AAK35694-X35747 represent cDNA sequences that encode novel polypeptides

XX (AAK02358-84) which are identified from a human placental cDNA library

XX by the signal sequence trap (SST) method. The polypeptides have a

XX broad range of physiological activity, including immunisation against

XX and inhibition of infections, allergies and cancer; regulation of tissue

XX formation and repair; activin/inhibin activity; chemokine/cytokine

XX activity; blood coagulation regulation; and receptor/ligand agonist

XX or antagonist activity. The polypeptides can be used for prevention

XX and treatment of disorders including infections by bacteria, yeasts and

XX viruses (including HIV) and protozoa; metabolic and hormonal disorders;

XX immune disorders (including severe combined immunodeficiency (SCID)

XX and AIDS; thrombosis; cancer; and traumatic or surgical wounds.

XX Sequence 1613 BP; 624 A; 243 C; 359 G; 387 T; 0 other:

Query Match 21.4%; Score 27.6; DB 20; Length 1613;  
 Best Local Similarity 52.6%; Pred. No. 6.4;  
 Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 7 tctctgtagcagattactcgagcattactcgagcacaagaaccccgagcagatgctcg 66  
 Db 729 TCTGCTGCATATTCATTCTTTCATATTTATTCGCTCAATGCTCACTGACTCTT 670  
 Oy 67 gggctttttgttggtgctgagctgtgtgccaacgcatattatccgagct 120  
 Db 669 GGTGTGCAATGATATTCGTGAGTCGAGGTCTCTCTGTTTATTTGTTTCACATT 616

## RESULT 6

AAV68588/c

ID AAV68588 standard; DNA: 2090 BP.

XX AAV68588;

XX 16-FEB-1999 (first entry)

XX Nucleotide sequence encoding the human nuclear protein.

XX Human nuclear protein; HEC; cell growth abnormality; hybridisation;

XX amplification; antibody; immunoassay; malignancy; mitosis; antigen;

XX immunoprecipitation; immunisation; vaccine; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 105..2033

XX /tag- a

XX /product= "human nuclear protein"

XX WO9845433-A1.

XX 15-OCT-1998.

XX 03-APR-1998; 98WO-US06727.

XX 04-APR-1997; 97US-0042609.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Alfred DC, Chamness GC, Clark GM, Hilsenbeck SG;

XX Osborn CK;

XX WPI: 1998-594481/50.

XX P-PSDB: AAW80950.

XX New nucleic acid encoding human highly expressed in cancer nuclear

XX protein - used for diagnosis and for modulation of the cell cycle to

XX control malignant and other cell growth abnormalities

XX Claim 1; Pages 55-58; 75pp; English.

XX This is the nucleotide sequence of a human nuclear protein (HEC) used

XX in the method of the invention to diagnose cell growth abnormalities.

XX Fragments of HEC are used to detect HEC-encoding nucleic acid, in

XX usual hybridisation or amplification assays, while antibodies are used

XX in conventional immunoassays to detect HEC or peptides. HEC is used to

XX modulate cell cycle progression (by disrupting chromatid separation)

XX and antibodies are used to disorder sister chromatid alignment and

XX separation in interphase cells, disrupting mitosis, particularly for

XX control of malignancy and other cell growth abnormalities at the

XX mitosis stage. Antibodies can also be used for protein purification, to

XX isolate sequences encoding HEC or related proteins; to study HEC

XX distribution in cells and to isolate antigens by immunoprecipitation.

XX HEC may also be used in vaccines and antibodies for passive

XX immunisation.

XX Sequence 2090 BP; 750 A; 362 C; 464 G; 514 T; 0 other:

Query Match 21.4%; Score 27.6; DB 19; Length 2090;  
 Best Local Similarity 52.6%; Pred. No. 6.9;  
 Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 7 tctctgtagcagattactcgagcattactcgagcacaagaaccccgagcagatgctcg 66  
 Db 1239 TCTGCTGCATATTCATTCTTTCATATTTATTCGCTCAATGCTCACTGACTCTT 1180  
 Oy 67 gggctttttgttggtgctgagctgtgtgccaacgcatattatccgagct 120  
 Db 1179 GGTGTGCAATGATATTCGTGAGTCGAGGTCTCTCTGTTTATTTGTTTCACATT 1126

## RESULT 7



|   |                          |
|---|--------------------------|
| AAV85564/C  | standard; cDNA; 2090 BP. |
| AAV38564  |                          |
| AAV38564;   |                          |
| 26-OCT-1998   | (first entry)            |
| Human nuclear protein HEC cDNA.                                       |                          |
| HCC gene; highly expressed in cancer; human; nuclear protein;         |                          |
| mitosis; cell cycle; cell proliferation; malignancy; vaccine; ss.     |                          |
| Homo sapiens.   |                          |
| Key   | Location/Qualifiers      |
| CDS   | 105..2033                |
|   | /tag=a                   |
| polyA_signal  | 2070..2075               |
|   | /tag=b                   |
| W09827994-A1.   |                          |
| 02-JUL-1998.  |                          |
| 18-DEC-1997;  | 97WO-US23385.            |
| 20-DEC-1996;  | 96US-0033600.            |
| (TEXA ) UNIV TEXAS SYSTEM.  |                          |
| Chen P, Chen Y, Lee W, Riley DJ;                                      |                          |
| WPI: 1998-377401/32.  |                          |
| P-PSDB; AAM62596.   |                          |
| Nucleic acid encoding human nuclear protein HEC modulating mitosis -  |                          |
| useful to, e.g control cell malignancies and other cell growth        |                          |
| abnormalities at mitosis stage and to produce HEC protein and         |                          |
| peptide(s)  |                          |
| Claim 2; Page 59-62; 93pp; English.                                   |                          |
| This cDNA clone codes for a novel human nuclear protein (see          |                          |
| AAM62596) designated HEC (highly expressed in cancer) that appears to |                          |
| be crucial for normal mitosis, possibly in regulation of normal       |                          |
| progression of M phase. A 1.8 kb cDNA fragment was originally         |                          |
| cloned from a human B cell cDNA library by its interaction with the   |                          |
| C-terminus of Rb in a yeast two-hybrid system. This fragment was      |                          |
| used as a probe to screen a human B cell cDNA library, and the        |                          |
| longest clone obtained included the full-length coding sequence for   |                          |
| HEC. HEC nucleic acid segments and their encoded polypeptides are     |                          |
| important in modulating mitosis and thus regulating cell              |                          |
| proliferation. They are useful in compositions to control cell        |                          |
| malignancies and other cell growth abnormalities at the cell          |                          |
| mitosis stage. DNA segments encoding HEC are also useful to           |                          |
| produce HEC protein/peptides and to detect HEC-encoding sequences     |                          |
| (e.g. from related species), by contacting with a nucleic acid        |                          |
| sample under suitable conditions and detecting hybridisation          |                          |
| (claimed kits are provided). HEC peptides are useful in vaccines      |                          |
| and may be administered to cells to disrupt chromatin separation      |                          |
| and so modulate cell cycle progression (claimed).                     |                          |
| Sequence 2090 BP; 750 A; 362 C; 464 G; 514 T; 0 other;                |                          |

| Query | Best Local Similarity   | Score                 | DB     | Length |
|-------|---|-----------------------|--------|--------|
| 1239  | 60; Conservative  | 21.48; Pred. No. 6.9; | 19;    | 2090;  |
|       | Matches   | 0;                    | Indels | Gaps   |
| 7     | tctctcgttagcagattacttcgagcattactgacgacgaagcccgacggagatgctcg       | 66                    |        |        |
|       |   |                       |        |        |
| Db    | tcctctgcgaattcatttttttttcattattattatgcgcacatgctctcaactcgatgactctt | 1180                  |        |        |

```
Oy      67 gggctccttgttgtygfygcfygacgyfllgtccaacgataattccgagct 120
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     1179 GGTGTCATGATATCTGTGATGTCGAGTCTTCTCCGTTTATTGTTTCACATT 1126
```

| RESULT     | 8                                |
|------------|----------------------------------|
| AAH81781/c |                                  |
| ID         | AAH81781 standard; DNA; 2150 BP. |

DT 21-SEP-2001 (first entry)

Human differential transcription-associated cDNA Seq ID 290.

xx Differential transcription; human; rat; tumour cell; cytostatic;  
 kw Ras modulator; Class II tumour suppressor gene; gene therapy; ss.  
 km

OS Homo sapiens.

PN W0200157058-A2.

PD 09-AUG-2001.

31-JAN-2001; 2001WO-EP01003.

PR 31-JAN-2000; 2000DE-1004102.  
XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
VY

P1 Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;  
PT Grins M, Hinzmann B, Schaefer R, Zuber J, Tchernitsa O;

Pl Grips M, Hellriegel M, Schmitz A, Sers C, XY

DR WPI; 2001-483415/52.

| PT | Nucleic acids differentially expressed between tumor and normal cells |
|----|---|
| PT | useful for diagnosis or therapy of tumors and for screening active    |
| PT | agents -  |

PS Disclosure; Page 444-445; 579pp; German.

This invention describes a nucleic acid (I) with differential expression between tumour and normal cells and which has cytostatic activity. (I) work as modulators of Ras activity by inducing expression of tumour suppressor genes. (I), and polypeptides encoded by them, are useful as targets for diagnosis or therapy and in screening to determine the effects of an active compound (potential pharmaceutical) on a cell line, particularly for diagnosis and treatment of tumours, especially by modulating expression of (I) (by gene therapy, antisense RNA or ribozyme methods) or by modulating the amount and/or location of (I)-encoded polypeptides (by administration of the polypeptide or its activator, antibody (optionally as a conjugate) or inhibitor). The method allows identification of many Class II tumour suppressor genes (i.e. genes that are not primary targets for tumour-initiating mutations). AAH81492-AAH82376 represent the human and rat derived nucleic acid fragments described in the method of the invention.

Sequence 2150 BP; 810 A; 362 C; 464 G; 514 T; 0 other;

|                       |                 |                |           |              |
|-----------------------|-----------------|----------------|-----------|--------------|
| Query Match           | 21.4%           | Score 27.6;    | DB 22;    | Length 2150; |
| Best Local Similarity | 52.6%;          | Pred. No. 7;   |           |              |
| Matches 60;           | Conservative 0; | Mismatches 54; | Indels 0; | Gap 0;       |

2y 7 tctgctcgtagcagattactctgcagacattactgacagacaagaaccccgacgagatgctcg 66  
 ||||| | ||||| | ||||| | ||||| | ||||| |  
 1239 TCTGCTGCATTCATTCTTTCATGATTTATTCGCTCAATGTCCTCACACTAGTACTTCT 1180

67 gggtcttttgttggtgctgtgacgtgtgtgtccaacgtattatccgact 120  
1179 GGTTCGCAATGATATTTCTGTAGTCGAGTGTCTCTCTGTTTATGTTCACATT 1126

| Query Match | Best Local Similarity | Score | DB         | Length     |
|-------------|-----------------------|-------|------------|------------|
| Matches     | 45; Conservative      | 0;    | Mismatches | 29; Indels |
|             |                       |       |            | Gaps       |
|             |                       |       |            | 0          |

2 gaccttcgtcgttagcagattactctcgagacataactcgaagacaagaccgcgacgagat 61

|   |      |  |      |
|---|------|--|------|
| Db  | 7636 | GACGACCGTCGCGGCGTATTTTTCGAGGTTTGTATCCGAAATAAAGCATCGAAG | 7637 |
| Qy  | 62   | gttcgaggctcttt   | 75   |
| Db  | 7636 | CATCGTGTGGT  | 7623 |
| RESULT  | 10   |  |      |
| AAAF21612   |      |  |      |
| AAAF21612   |      | standard; DNA; 349980 BP.                              |      |
| AAAF21612;  |      |  |      |
| 13-MAR-2001   |      | (first entry)  |      |
| Neisseria meningitidis B  |      | nucleotide sequence SEQ ID NO:113.                     |      |
| Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;      |      |  |      |
| diagnosis; antigen; detection; infection; gene therapy; antibacterial;    |      |  |      |
| ds.   |      |  |      |
| Neisseria meningitidis.   |      |  |      |
| WO200066791-A1.   |      |  |      |
| 09-NOV-2000.  |      |  |      |
| 08-MAR-2000; 2000WO-US05928.  |      |  |      |
| 30-APR-1999; 99US-0132068.  |      |  |      |
| 08-OCT-1999; 99WO-US23573.  |      |  |      |
| 28-FEB-2000; 2000GB-0004695.  |      |  |      |
| (CHIR ) CHIRON CORP.  |      |  |      |
| (GENO ) INST GENOMIC RES.   |      |  |      |
| Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V,         |      |  |      |
| Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R,         |      |  |      |
| Frazer CM, Grandi G.  |      |  |      |
| WPI; 2000-647603/62.  |      |  |      |
| Neisseria meningitidis B  |      | full length genome sequence and open reading           |      |
| frames are used to detect, treat and prevent Neisserial infections -      |      |  |      |
| Claim 7; Appendix A; 692pp; English.                                      |      |  |      |
| The present invention describes the full length genome of                 |      |  |      |
| Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607    |      |  |      |
| to AAF21613 represent fragments of the NMB genomic sequence, as the       |      |  |      |
| sequence was too long to go in a record on its own it was split into 8    |      |  |      |
| sequences which overlap each other at the beginning and end of each       |      |  |      |
| sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at   |      |  |      |
| the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at  |      |  |      |
| the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the    |      |  |      |
| Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to         |      |  |      |
| AAF21606 represent PCR primers which are used in the exemplification of   |      |  |      |
| the present invention. The NMB genome and fragments from it have          |      |  |      |
| antibacterial activity, and can be used in vaccines and gene therapy.     |      |  |      |
| Neisseria nucleic acids, proteins and/or antibodies which binds to the    |      |  |      |
| proteins can be used in compositions for treating or preventing infection |      |  |      |
| due to Neisserial bacteria or as a diagnostic reagent for detecting the   |      |  |      |
| presence of Neisserial bacteria or of antibodies raised to Neisserial     |      |  |      |
| bacteria. Computers, computer memory, computer storage medium or computer |      |  |      |
| databases can be used in a search to identify open reading frames (ORFs)  |      |  |      |
| or coding sequences within the NMB genome. The DNA sequences provide      |      |  |      |
| further opportunities to find antigenic or immunogenic proteins which are |      |  |      |
| more effective in vaccines than the outer membrane proteins currently     |      |  |      |
| used.   |      |  |      |
| Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;          |      |  |      |



|                       |   |               |
|-----------------------|---|---------------|
| PR                    | 31-JUL-1998;  | 98US-0094935. |
| PR                    | 10-AUG-1998;  | 98US-0095880. |
| PR                    | 11-AUG-1998;  | 98US-0096068. |
| XX                    | 06-MAY-1999;  | 99US-0096068. |
| PA                    | (GENY ) GENETICS INST INC.  |               |
| PI                    | Jacobs K, MCCOY JM, LAVALLIE ER, COLLINS-RACLE LA, EVANS C;               |               |
| PI                    | MERBERG D, TREACY M, AGOSTINO MJ, STEININGER RJ, BOWMAN MR;               |               |
| PI                    | DIBLASIO-SMITH E, WIDOM A;  |               |
| XX                    |   |               |
| DR                    | WPI; 2000-052937/04.  |               |
| DR                    | P-PSDB; AAU53015.   |               |
| PT                    | New polynucleotides encoding secreted human proteins, derived from        |               |
| PT                    | adult placenta, adult retina, fetal brain, fetal                          |               |
| PS                    | Claim 44; Page 395-396; 492pp; English.                                   |               |
| XX                    |   |               |
| CC                    | The present invention describes new human secreted proteins which were    |               |
| CC                    | isolated from adult placenta, adult retina, foetal brain, foetal kidney,  |               |
| CC                    | adult blood, adult brain, adult thyroid, adult bladder, adult neural      |               |
| CC                    | tissue, adult testes, and adult lymph node cDNA libraries. The human      |               |
| CC                    | secreted proteins, and the polynucleotides encoding them, are predicted   |               |
| CC                    | to have biological activities which would make them suitable for          |               |
| CC                    | treating, preventing or ameliorating medical conditions in humans and     |               |
| CC                    | animals. Suggested activities include nutritional activity, cytokine      |               |
| CC                    | and cell proliferation/differentiation activity, immune stimulating       |               |
| CC                    | (e.g. as vaccines) or suppressing activity, haematopoiesis regulating     |               |
| CC                    | activity, tissue growth activity, activin/inhibin activity,               |               |
| CC                    | chemoattractic/chemokinetic activity, hemostatic and thrombolytic         |               |
| CC                    | activity, receptor/ligand activity, anti-inflammatory activity,           |               |
| CC                    | cadherin/tumour invasion suppressor activity, and tumour inhibition       |               |
| CC                    | activity. The polynucleotides are also stated to be useful for gene       |               |
| CC                    | therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and         |               |
| CC                    | AAV52998 to AAV53060 represent human secreted proteins, given in the      |               |
| XX                    | present invention.  |               |
| XQ                    |   |               |
| Sequence              | 1670 BP; 362 A; 447 C; 472 G; 389 T; 0 other;                             |               |
| Query Match           | 21.2%; Score 27.4; DB 21; Length 1670;                                    |               |
| Best Local Similarity | 54.5%; Pred. No. 7.6;   |               |
| Matches               | 55; Conservative 0; Mismatches 46; Indels 0; Gaps 0                       |               |
| OY                    | 24 ctctgagattactcgaagaaccgcgcagacagatgtcgggcttttcttggtg 83                |               |
| DB                    | 493 ctctgaggattcttcgttgattaacctccacccctctggcgagtgtttcttg 552              |               |
| OY                    | 84 tctgttagtgtgttcacaacgcatattctcggaacttt 124                             |               |
| DB                    | 553 ggcctgaacctgcgcacattcaatcaatgacctactcgt 593                           |               |
| RESULT 13             |   |               |
| ID                    | AAH16020  |               |
| ID                    | AAH16020 standard; cDNA; 1878 BP.   |               |
| AC                    | AAH16020;   |               |
| DT                    | 26-JUN-2001 (first entry)   |               |
| DE                    | Human cDNA sequence SEQ ID NO:14672.                                      |               |
| HM                    | Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss. |               |
| OS                    | Homo sapiens.   |               |
| PN                    | EP1074617-A2.   |               |
| PD                    | 07-FEB-2001.  |               |
| PF                    | 28-JUL-2000; 2000EP-0116126.  |               |

|        |   |   |
|--------|---|---|
| XX     | 29-JUL-1999;  | 99JP-0248036.                                     |
| PR     | 27-AUG-1999;  | 99JP-0300253.                                     |
| PR     | 11-JAN-2000;  | 2000JP-0118776.                                   |
| PR     | 02-MAY-2000;  | 2000JP-0183767.                                   |
| PR     | 09-JUN-2000;  | 2000JP-0241899.                                   |
| XX     |   |   |
| PA     | (HELI-) HELIX RES INST.   |   |
| PI     | Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;             |   |
| EI     | Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;                      |   |
| XX     |   |   |
| DR     | WPI; 2001-318749/34.  |   |
| XX     |   |   |
| PT     | Primer sets for synthesizing polynucleotides, particularly the 5602       |   |
| PT     | full-length CDNA's defined in the specification, and for the detection    |   |
| PT     | and/or diagnosis of the abnormality of the proteins encoded by the        |   |
| PT     | full-length CDNA's -  |   |
| PS     | Claim 8; SEQ ID 14672; 2537pp + CD ROM; English.                          |   |
| XX     |   |   |
| CC     | The present invention describes primer sets for synthesising 5602         |   |
| CC     | full-length CDNA's defined in the specification. Where a primer set       |   |
| CC     | comprises: (a) an oligo-dT primer and an oligonucleotide complementary    |   |
| CC     | to the complementary strand of a polynucleotide which comprises one of    |   |
| CC     | the 5602 nucleotide sequences defined in the specification, where the     |   |
| CC     | oligonucleotide comprises at least 15 nucleotides; or (b) a combination   |   |
| CC     | of an oligonucleotide comprising a sequence complementary to the          |   |
| CC     | complementary strand of a polynucleotide which comprises a 5'-end         |   |
| CC     | sequence and an oligonucleotide comprising a sequence complementary to a  |   |
| CC     | polynucleotide which comprises a 3'-end sequence, where the               |   |
| CC     | oligonucleotide comprises at least 15 nucleotides and the combination of  |   |
| CC     | the 5'-end sequence/3'-end sequence is selected from those defined in     |   |
| CC     | the specification. The primer sets can be used in antisense therapy and   |   |
| CC     | in gene therapy. The primers are useful for synthesising polynucleotides, |   |
| CC     | particularly full-length CDNA's. The primers are also useful for the      |   |
| CC     | detection and/or diagnosis of the abnormality of the proteins encoded by  |   |
| CC     | the full-length CDNA's. The primers allow obtaining of the full-length    |   |
| CC     | CDNA's easily without any specialised methods. AAH03166 to AAH13628 and   |   |
| CC     | AAH13633 to AAH18742 represent human cDNA sequences; AAPB2446 to          |   |
| CC     | AAPB9593 represent human amino acid sequences; and AAH13629 to AAH13632   |   |
| CC     | represent oligonucleotides, all of which are used in the exemplification  |   |
| CC     | of the present invention.   |   |
| XX     |   |   |
| SO     | Sequence 1878 BP; 356 A; 526 C; 567 G; 429 T; 0 other:                    |   |
|        |   |   |
|        | Query Match   | 21.2%; Score 27.4; DB 22; Length 1878;            |
|        | Best Local Similarity   | 54.3%; Prid. No. 7.8;                             |
|        | Matches   | 55; Conservative 0; Mismatches 46; Indels 0; Gaps |
| OY     | 24 ctctgagattacgcagaacaagaccggcagcgatgtcgaggcttttgttgsg                   | 83  |
| DB     | 743   |   |
|        | ctctgggagtcttcgttggattaacctccaccactccctcgggcgatgtgtttcttgsg               | 802   |
| OY     | 84 tgcctgacgctgtgtccaacgcgatattatccggagtagt                               | 124   |
|        |   |   |
| DB     | 803 ggcctgaacctgtgcgcccaatcatcatgctaactgt                                 | 843   |
|        |   |   |
| RESULT | 14  |   |
| ID     | AAAX37414   |   |
| XX     | AAAX37414 standard; CDNA; 1951 BP.  |   |
| AC     | AAAX37414:  |   |
| DT     | 06-JUL-1999 (first entry)   |   |
| DE     | Human secreted protein cDNA fragment containing gene 46.                  |   |
| KM     | Human; secreted protein; prevention; treatment; protein therapy;          |   |
| KM     | gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;      |   |
| KM     | developmental abnormality; foetal deficiency; blood disorder; lymphoma;   |   |

|                       |  |  |
|-----------------------|--|--|
| KW                    |  | leukemia; immune system disorder; autoimmune disease; hepatic disease; |
| KV                    |  | renal disease; inflammation; allergy; asthma; sepsis; diabetes; AIDS;  |
| KM                    |  | Alzheimer's disease; cognitive disorder; schizophrenia; osteoporosis;  |
| XX                    |  | arthritis; psoriasis; digestive; endocrine; infection; ss.             |
| OS                    | Homo sapiens.  |  |
| PN                    | M09090155-A1.  |  |
| PD                    | 25-FEB-1999.   |  |
| PF                    | 18-AUG-1998;   | 98MO-US17044.  |
| PR                    | 16-JUN-1998;   | 98US-0092956.  |
| PR                    | 13-JUL-1998;   | 98US-0092956.  |
| PR                    | 19-AUG-1997;   | 97US-0056368.  |
| PR                    | 19-AUG-1997;   | 97US-0056369.  |
| PR                    | 19-AUG-1997;   | 97US-0056535.  |
| PR                    | 19-AUG-1997;   | 97US-0056535.  |
| PR                    | 19-AUG-1997;   | 97US-0056556.  |
| PR                    | 19-AUG-1997;   | 97US-0056628.  |
| PR                    | 19-AUG-1997;   | 97US-0056629.  |
| PR                    | 19-AUG-1997;   | 97US-0056726.  |
| XX                    | 19-AUG-1997;   | 97US-0056728.  |
| PA                    | (HUMA-) HUMAN GENOME SCI INC.  |  |
| PI                    | Brewer LA, Duan R, Ebner R, Endress GA,                                  | Feng P;  |
| PI                    | Florence CA, Florence KA, Komatsoulis SA,                                | Lafleur DW;  |
| PI                    | Moore PA, Olsen HS, Rosen CA, Ruben SW,                                  | Shi Y, Soppet DR;  |
| PI                    | Young PE.  |  |
| DR                    | WPI, 1999-190160/16.   |  |
| DR                    | P-PSDB: AAY07789.  |  |
| PT                    | New isolated human genes and the secreted polypeptides they encode       |  |
| PT                    | - useful for diagnosis and treatment of e.g. cancers, neurological       |  |
| XX                    | disorders, immune diseases, inflammation or blood disorders              |  |
| PS                    | Claim 1a; Page 216-217; 280pp; English.                                  |  |
| CC                    | This invention describes novel isolated human secreted proteins and      |  |
| CC                    | their encoding nucleic acid sequences. The products of the invention     |  |
| CC                    | are useful for preventing, treating or ameliorating medical conditions   |  |
| CC                    | e.g. by protein or gene therapy. Also pathological conditions can be     |  |
| CC                    | diagnosed by determining the presence or amount of expression of         |  |
| CC                    | the new polypeptides in a sample or by determining the presence or       |  |
| CC                    | absence of mutations in the new polynucleotides. Specific uses are       |  |
| CC                    | described for each of the 70 polynucleotides, based on which tissues     |  |
| CC                    | they are most highly expressed in, and include developing products for   |  |
| CC                    | the diagnosis or treatment of cancer, tumours, neurodegenerative         |  |
| CC                    | disorders, developmental abnormalities and foetal deficiencies, blood    |  |
| CC                    | diseases, hepatic and renal disease, lymphomas, inflammation, allergies, |  |
| CC                    | asthma, sepsis, diabetes, Alzheimer's and cognitive disorders,           |  |
| CC                    | schizophrenia, osteoporosis, arthritis, psoriasis, digestive/endocrine   |  |
| CC                    | disorders, infections and AIDS. The human secreted proteins of the       |  |
| CC                    | invention are represented in AAY07744-Y07850 and the encoding nucleic    |  |
| CC                    | acids are represented in AAX37369-X37441.                                |  |
| XX                    |  |  |
| XX                    |  |  |
| Sequence              | 1951 BP; 403 A; 536 C; 576 G; 431 T; 5 other;                            |  |
| Query Match           | 21.2%; Score 27.4; DB 20; Length 1951;                                   |  |
| Best Local Similarity | 54.5%; Pred. No. 7.9;  |  |
| Matches               | 55; Conservative 0; Mismatches 46; Indels 0; Gaps 0;                     |  |
| OY                    | 24 cttagagcattactgaagacaagaaccgcgagatgtctgggttttcttggtgg 83              |  |
|                       |  |  |
| Db                    | 718 ctgcctgagcatcttcgttgattaatccctcaaccctccggcgagatgtttcttggtg 777       |  |
|                       |  |  |
| OY                    | 84 tgcctgagcgtgtgtccaaccgcttatcttcgcgactagt 124                          |  |
|                       |  |  |

Db 778 ggcgttaaccgcgcgcacattcataatgacctactgt 818

RESULT 15

AAH01309 standard; DNA; 2628 BP.

AAH01309;

24-JUL-2001 (first entry)

Escherichia coli nucleotide sequence SEQ ID NO:1300.

Species specific; genus specific; family specific; probe; detection; identification; algal; archaeal; bacterial; fungal; parasitical; microorganism; diagnosis; translation elongation factor Tu; toxin; translation elongation factor G; RecA recombinase; resistance; catalytic subunit of proton-translocating ATPase; antimicrobial; vaccine; primer; ds.

Escherichia coli.

WO200123604-A2.

05-APR-2001.

28-SEP-2000; 2000MO-CA01150.

28-SEP-1999; 99CA-2283458.

19-MAY-2000; 2000CA-2307010.

(INFE-) INFECTIO DIAGNOSTIC (IDI) INC.

Bergeron MG, Bollssnot M, Huletsky A, Menard C, Ouellette M; Picard FJ, Roy PH;

WPI: 2001-245006/25.

Nucleic acid sequences are used to generate universal probes and primers which can be used to identify and detect the presence of algal, archaeal, bacterial, fungal and parasitical species in a test sample -

Disclosure: Page 1109-1110; 1580pp; English.

The present invention describes a method for generating a repertoire of nucleic acids of tuf, fus, atpD and/or recA genes from which probes and/or primers are derived. The method comprises amplifying the nucleic acids of determined algal, archaeal, bacterial, fungal and parasitical species with a combination of defined primer pairs. The method can be used for producing probes and/or primers for detecting one or more related microorganisms e.g. algae, archaea, bacteria, fungi and parasites, for universal detection and for specific and ubiquitous detection and identification of an algal, archaeal, bacterial, fungal and parasitical species, genus, family and group. A nucleic acid (I) obtained using the method of the invention can be used for the universal detection of any bacterium, fungus or parasite in a sample and for the detection of at least one antimicrobial agent resistance gene or at least one toxin gene. hexa nucleic acids are used for the specific and ubiquitous detection and for identification of Streptococcus pneumoniae microorganisms. Microbial species or genus or family or phylum or group which can be detected include Abiotrophia adiacens, Bordetella sp., Corynebacterium sp., Enterobacteriaceae group, Escherichia coli, Mycobacteriaceae family, Pseudomonas group, Streptococcus sp., Neisseria gonorrhoeae and Staphylococcus sp. Using DNA based tests provides faster results than substrate specificity tests as results can be determined in an hour and improved accuracy is also achieved. AAH00010 to AAH00204 represent nucleotide sequences and primers/probes are given in the exemplification of the present invention.

Sequence 2628 BP; 646 A; 662 C; 756 G; 564 T; 0 other;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:10:40 ; Search time 112.49 Seconds  
(without alignments)  
259.718 Million cell updates/sec

Title: US-09-816-391A-1\_COPY\_472\_600  
Perfect score: 129  
Sequence: 1 tgcattctctgctgtagcga.....tattcgagctagttacgcg 129

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued\_Patents\_MA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID                | Description       |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1          | 28.4  | 22.0        | 936    | 2 US-08-602-359A-31  | Sequence 31, Appl |
| 2          | 26.2  | 20.3        | 907    | 3 US-08-672-850-7    | Sequence 7, Appl  |
| 3          | 26.2  | 20.3        | 907    | 3 US-08-672-850-11   | Sequence 11, Appl |
| 4          | 26.2  | 20.3        | 1001   | 3 US-08-672-850-10   | Sequence 10, Appl |
| 5          | 26    | 20.2        | 43280  | 2 US-08-804-227C-1   | Sequence 1, Appl  |
| 6          | 25.8  | 20.0        | 1158   | 4 US-09-198-092-1    | Sequence 1, Appl  |
| 7          | 25.8  | 20.0        | 2262   | 2 US-08-674-887A-5   | Sequence 5, Appl  |
| 8          | 25.8  | 20.0        | 2262   | 2 US-08-951-844-5    | Sequence 5, Appl  |
| 9          | 25.4  | 19.7        | 1586   | 1 US-08-461-244-1    | Sequence 1, Appl  |
| 10         | 25.2  | 19.5        | 50341  | 1 US-08-247-901C-1   | Sequence 1, Appl  |
| 11         | 25.2  | 19.5        | 50341  | 2 US-09-075-904-1    | Sequence 1, Appl  |
| 12         | 25.2  | 19.5        | 52297  | 4 US-09-426-436-1    | Sequence 1, Appl  |
| 13         | 25.2  | 19.5        | 52297  | 4 US-08-705-557-1    | Sequence 1, Appl  |
| 14         | 25    | 19.4        | 5799   | 1 US-08-483-101-1    | Sequence 1, Appl  |
| 15         | 24.8  | 19.2        | 799    | 1 US-08-173-510B-100 | Sequence 100, App |
| 16         | 24.8  | 19.2        | 799    | 1 US-08-456-218-98   | Sequence 98, App  |
| 17         | 24.8  | 19.2        | 799    | 2 US-08-450-497-100  | Sequence 100, App |
| 18         | 24.2  | 18.8        | 1983   | 4 US-08-991-677-3    | Sequence 3, Appl  |
| 19         | 23.8  | 18.4        | 2326   | 2 US-08-231-193A-41  | Sequence 41, Appl |
| 20         | 23.8  | 18.4        | 2326   | 2 US-08-486-273A-41  | Sequence 41, Appl |
| 21         | 23.8  | 18.4        | 2326   | 3 US-08-480-474-41   | Sequence 41, Appl |
| 22         | 23.8  | 18.4        | 2326   | 3 US-08-486-273A-41  | Sequence 41, Appl |
| 23         | 23.8  | 18.4        | 3243   | 2 US-08-231-193A-44  | Sequence 44, Appl |
| 24         | 23.8  | 18.4        | 3243   | 2 US-08-480-474-44   | Sequence 44, Appl |
| 25         | 23.8  | 18.4        | 3243   | 2 US-08-486-273A-44  | Sequence 44, Appl |
| 26         | 23.8  | 18.4        | 3243   | 3 US-08-940-086A-44  | Sequence 44, Appl |
| 27         | 23.8  | 18.4        | 3698   | 2 US-08-231-193A-43  | Sequence 43, Appl |

|    |      |      |      |                     |                   |
|----|------|------|------|---------------------|-------------------|
| 28 | 23.8 | 18.4 | 3698 | 2 US-08-486-273A-43 | Sequence 43, Appl |
| 29 | 23.8 | 18.4 | 3698 | 3 US-08-480-474-43  | Sequence 43, Appl |
| 30 | 23.8 | 18.4 | 3698 | 3 US-08-940-086A-43 | Sequence 43, Appl |
| 31 | 23.8 | 18.4 | 4002 | 2 US-08-231-193A-53 | Sequence 53, Appl |
| 32 | 23.8 | 18.4 | 4002 | 2 US-08-486-273A-53 | Sequence 53, Appl |
| 33 | 23.8 | 18.4 | 4002 | 3 US-08-480-474-53  | Sequence 53, Appl |
| 34 | 23.8 | 18.4 | 4002 | 3 US-08-480-474-53  | Sequence 53, Appl |
| 35 | 23.8 | 18.4 | 4017 | 2 US-08-231-193A-49 | Sequence 49, Appl |
| 36 | 23.8 | 18.4 | 4017 | 2 US-08-486-273A-49 | Sequence 49, Appl |
| 37 | 23.8 | 18.4 | 4017 | 2 US-08-480-474-49  | Sequence 49, Appl |
| 38 | 23.8 | 18.4 | 4017 | 3 US-08-940-086A-49 | Sequence 49, Appl |
| 39 | 23.8 | 18.4 | 4053 | 2 US-08-231-193A-47 | Sequence 47, Appl |
| 40 | 23.8 | 18.4 | 4053 | 2 US-08-486-273A-47 | Sequence 47, Appl |
| 41 | 23.8 | 18.4 | 4053 | 3 US-08-480-474-47  | Sequence 47, Appl |
| 42 | 23.8 | 18.4 | 4053 | 3 US-08-940-086A-47 | Sequence 47, Appl |
| 43 | 23.8 | 18.4 | 4068 | 2 US-08-231-193A-5  | Sequence 5, Appl  |
| 44 | 23.8 | 18.4 | 4068 | 2 US-08-486-273A-5  | Sequence 5, Appl  |
| 45 | 23.8 | 18.4 | 4068 | 3 US-08-480-474-5   | Sequence 5, Appl  |

#### ALIGNMENTS

RESULT 1  
US-08-602-359A-31  
Sequence 31, Application US/08602359A  
Patent No. 5942430  
GENERAL INFORMATION:  
APPLICANT: ROBERTSON, Daniel E.  
APPLICANT: MURPHY, Dennis  
APPLICANT: REID, John  
APPLICANT: MAFRA, Anthony  
APPLICANT: LINK, Steven  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: KOSMOTKA, Anna  
TITLE OF INVENTION: ESTERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & RICHARDSON P.C.  
STREET: 4225 EXECUTIVE SQUARE, STE 1400  
CITY: LA JOLLA  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,359A  
FILING DATE: February 16, 1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09010/010001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 936 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
US-08-602-359A-31





TOPOLOGY: 1linear  
MOLECULE TYPE: DNA (genomic)

RESULT 7  
US-08-674-887A-5/c

Sequence 5, Application US/08674887A  
Patent No. 5939300  
GENERAL INFORMATION:  
APPLICANT: Robertson, Dan E.  
APPLICANT: Sanyal, Indrajit  
APPLICANT: Adhikari, Robert S.  
TITLE OF INVENTION: CATALASES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674,887A  
FILING DATE: 03-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Haile, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09015/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2262 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...2259  
US-08-674-887A-5

Query Match 20.0%; Score 25.8; DB 2; Length 2262;  
Best Local Similarity 60.9%; Pred. No. 4.4;  
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

DB 13 cgtagcgattactctgagcattactgagcaagaagaccgagcgagatggtcggtct 72  
916 CATTGCCATTACCGTACCATTTACCGAGCGGTGGCCGCCAGCTGTGAGGCTGCGGTTT 857

QY 73 ttgtgtgt 81  
||| ||| |||  
DB 856 TTTCGTCTGT 848

RESULT 8  
US-08-951-844-5/c  
Sequence 5, Application US/08951844  
Patent No. 6074860  
GENERAL INFORMATION:  
APPLICANT: Robertson et al.  
TITLE OF INVENTION: Catalases  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY

COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/951,844  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/674,887  
FILING DATE: JULY 3, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Charles J. Heron  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 331400-55  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2262 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: Oligonucleotide  
US-08-951-844-5

Query Match 20.0%; Score 25.8; DB 3; Length 2262;  
Best Local Similarity 60.9%; Pred. No. 4.4;  
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

DB 13 cgtagcgattactctgagcattactgagcaagaagaccgagcgagatggtcggtct 72  
916 CATTGCCATTACCGTACCATTTACCGAGCGGTGGCCGCCAGCTGTGAGGCTGCGGTTT 857

QY 73 ttgtgtgt 81  
||| ||| |||  
DB 856 TTTCGTCTGT 848

RESULT 9  
US-08-461-244-1/c  
Sequence 1, Application US/08461244  
Patent No. 5776729  
GENERAL INFORMATION:  
APPLICANT: Soppet, Daniel R.  
APPLICANT: Yl, Li  
APPLICANT: Ruben, Steven M.  
APPLICANT: Rosen, Craig A.  
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HGBER32  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,  
ADDRESSEE: STUART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,244  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

LENGTH: 5041  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE:  
 DESCRIPTION: U5 shuttle phasmid sequence  
 HYDROTHERMICAL: NO  
 ANTI-SENSE:

ZAF: 1001B  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 Inch 1.44 Mb storage diskette  
 COMPUTER: IBM PC Compatible  
 OPERATING SYSTEM: MS-DOS  
 SOFTWARE: Word Processor (ASCII)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/075,904  
 FILING DATE: May 11, 1998  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/247,901  
 FILING DATE: May 23, 1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bogosian, Elizabeth A  
 REGISTRATION NUMBER: 39,911  
 REFERENCE/DOCKET NUMBER: 96700/475

```

? ZIP: 10016
? COMPILER READABLE FORM:
? MEDIUM TYPE: 3.5 Inch 1.44 MB storage diskette
? COMPUTER: IBM PC Compatible
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PROCESSOR (ASCII)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/426,436
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/705,557
? FILING DATE:
? APPLICATION NUMBER: US/08/057,531
? FILING DATE:
? APPLICATION NUMBER: 07/833,431
? FILING DATE: February 7, 1992
? ATTORNEY/AGENT INFORMATION:
? NAME: Pasqualini, Patricia A.
? REGISTRATION NUMBER: 34,894
? REFERENCE/DOCKET NUMBER: 96700/228
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 697-5995
? TELEFAX: (212) 286-0854 or 286-0082
? TELEX: TWX 710-581-4766
? INFORMATION FOR SEO ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 52297
? TYPE: nucleotide
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE:
? DESCRIPTION: phage genome sequence
? HYPOTHETICAL: no
? ANTI-SENSE: no
? FRAGMENT TYPE: not applicable.
? ORIGINAL SOURCE:
? ORGANISM: mycobacteriophage L5
? STRAIN: not applicable
? INDIVIDUAL ISOLATE: L5
? DEVELOPMENTAL STAGE: not applicable
? HAPOTYPE: not applicable
? TISSUE TYPE: not applicable
? CELL TYPE: not applicable
? CELL LINE: not applicable
? ORGANELLE: not applicable
? IMMEDIATE SOURCE: mycobacteriophage L5 particles
? POSITION IN GENOME: entire genome
? FEATURE:
? NAME/KEY:
? LOCATION:
? IDENTIFICATION METHOD:
? OTHER INFORMATION:
? PUBLICATION INFORMATION:
? AUTHORS: Hatfull and Sarkis
? TITLE: DNA Sequence, Structure and Gene
? TITLE: Expression of Mycobacteriophage L5:
? TITLE: A Phage System for Mycobacterial
? TITLE: Genetics
? JOURNAL: Molecular Microbiology
? VOLUME: 7
? PAGES: 395-405
? DATE: 1993
? US-09-426-436-1
?
? Query Match 19.5%; Score 25.2; DB 4; Length 52297;
? Best Local Similarity 66.7%; Pred.No.26;
? Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0

```

RESULT 13  
US-08-705-557-1  
Sequence 1, Application US/08705557  
Patent No. 6300061  
GENERAL INFORMATION:  
APPLICANT: William R. Jacobs, Jr.  
APPLICANT: Barry R. Bloom  
APPLICANT: Graham F. Hatfull  
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC  
NUMBER OF SEQUENCES: 1  
REPORTER MYCOBACTERIOPHAGES  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Edenstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Word Processor (ASCII)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/705,557  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/057,531  
FILING DATE:  
APPLICATION NUMBER: 07/833,431  
FILING DATE: February 7, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pasqualini, Patricia A.  
REGISTRATION NUMBER: 34,894  
REFERENCE/DOCKET NUMBER: 96700/238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 697-5995  
TELEFAX: (212) 286-0854 or 286-0082  
TELEX: TWX 710-581-4766  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 52297  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: phage genome sequence  
HYPOTHETICAL: no  
ANTI-SENSE: no  
FRAGMENT TYPE: not applicable.  
ORIGINAL SOURCE:  
ORGANISM: mycobacteriophage L5  
STRAIN: not applicable  
INDIVIDUAL ISOLATE: L5  
DEVELOPMENTAL STAGE: not applicable  
HAPLOTYPE: not applicable  
TISSUE TYPE: not applicable  
CELL TYPE: not applicable  
CELL LINE: not applicable  
ORGANELLE: not applicable  
IMMEDIATE SOURCE: mycobacteriophage L5 particles  
POSITION IN GENOME: entire genome  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS: Hatfull and Sarkis  
TITLE: DNA Sequence, Structure and Gene

;; TITLE: Expression of Mycobacteriophage L5:  
;; TITLE: A Phage System for Mycobacterial  
;; TITLE: Genetics  
;; JOURNAL: Molecular Microbiology  
;; VOLUME: 7  
;; PAGES: 395-405  
;; DATE: 1993  
US-08-705-557-1  
Query Match 19.5%; Score 25.2; DB 4; Length 52297;  
Best Local Similarity 66.7%; Pred. No. 26;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;  
OY 54 accggatggtggtgggtcttcttgggtggtgctgacgtgtgtccacgct 107  
Db. 1773 ACCGTGACCATCGGCGACATCGCTGATGCTGACGCTGGACGTAATCCCGCGCGT 1826  
RESULT 14  
US-08-483-101-1  
Sequence 1, Application US/08483101  
Patent No. 5932715  
GENERAL INFORMATION:  
APPLICANT: Scott, June R.  
APPLICANT: Froehlich, Barbara  
APPLICANT: Caron, Judy  
TITLE OF INVENTION: CS2 Proteins and Coding Sequences  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Wanner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,101  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 6-95  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5798 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 499..1215  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 499..552  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 553..1212  
FEATURE:





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 14:42:22 ; Search time 2300.01 Seconds  
(without alignments)  
602.695 Million cell updates/sec

Title: US-09-816-391a-1\_COPY\_472\_600  
Perfect score: 129  
Sequence: 1 tgacctctgctgctagcga.....tattcgagactagtcagcg 129

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues  
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_estl2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pin:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 31.8  | 24.7        | 600    | 11    | BG817831 ESSU0253  |
| 2          | 30.8  | 23.9        | 524    | 10    | AA517086 vH98110.r |
| 3          | 30.6  | 23.7        | 199    | 10    | AJ282728 4A3A-AB-  |
| 4          | 30.6  | 23.7        | 677    | 10    | AV745983 AV745983  |
| 5          | 30.2  | 23.4        | 546    | 11    | BG817800 ESSU0222  |
| 6          | 30.2  | 23.4        | 577    | 11    | BG817800 ESSU0222  |
| 7          | 30.2  | 23.4        | 599    | 11    | BG817905 ESSU0327  |
| 8          | 30.2  | 23.4        | 657    | 11    | BG817854 ESSU0276  |
| 9          | 30.2  | 23.4        | 682    | 11    | BG817862 ESSU0284  |
| 10         | 30.2  | 23.4        | 720    | 11    | BG817630 ESSU0052  |
| 11         | 30.2  | 23.4        | 1101   | 13    | CNS0170C           |
| 12         | 29.6  | 22.9        | 187    | 11    | BF841751 BF841751  |

|    |      |      |      |    |          |                     |
|----|------|------|------|----|----------|---------------------|
| 13 | 29.6 | 22.9 | 903  | 13 | CNS04GSE | AL290039 Tetradon   |
| 14 | 29.4 | 22.8 | 485  | 11 | H76502   | H76502 18207 Lambd  |
| 15 | 29.4 | 22.8 | 1353 | 11 | B1114606 | B1114606 602861984  |
| 16 | 29.2 | 22.6 | 259  | 10 | A1613789 | A1613789 vH98110.y  |
| 17 | 29.2 | 22.6 | 345  | 11 | BE945901 | BE945901 UI-M-B20-  |
| 18 | 29.2 | 22.6 | 365  | 11 | W48020   | W48020 mc91e03.r1   |
| 19 | 29.2 | 22.6 | 406  | 10 | BE651993 | BE651993 UI-M-APO-  |
| 20 | 29.2 | 22.6 | 415  | 10 | AA103400 | AA103400 mo24f02.r  |
| 21 | 29.2 | 22.6 | 420  | 10 | BE226847 | BE226847 us73d03.y  |
| 22 | 29.2 | 22.6 | 534  | 11 | BE914799 | BE914799 60167823   |
| 23 | 29.2 | 22.6 | 653  | 11 | BG965887 | BG965887 602829669  |
| 24 | 29.2 | 22.6 | 695  | 11 | B1079667 | B1079667 602827635  |
| 25 | 29.2 | 22.6 | 763  | 11 | BG922669 | BG922669 602821635  |
| 26 | 29.2 | 22.6 | 764  | 11 | BF384357 | BF384357 602046710  |
| 27 | 29.2 | 22.6 | 790  | 10 | AU080797 | AU080797 AU080797   |
| 28 | 29.2 | 22.6 | 791  | 11 | BF118685 | BF118685 601754973  |
| 29 | 29.2 | 22.6 | 838  | 11 | BF247036 | BF247036 602359684  |
| 30 | 29.2 | 22.6 | 848  | 11 | BE532038 | BE532038 602073004  |
| 31 | 29.2 | 22.6 | 850  | 11 | BF781560 | BF781560 602104468  |
| 32 | 29.2 | 22.6 | 863  | 11 | BF784104 | BF784104 602110059  |
| 33 | 29.2 | 22.6 | 908  | 11 | B1248625 | B1248625 602993062  |
| 34 | 29.2 | 22.6 | 908  | 11 | AA499104 | AA499104 v192c10.r  |
| 35 | 29.2 | 22.5 | 940  | 13 | A0572855 | A0572855 HS_-5398_B |
| 36 | 29.2 | 22.5 | 533  | 11 | BG001371 | BG001371 OVA-GN014  |
| 37 | 29.2 | 22.5 | 1019 | 13 | CNS05L5T | AL342362 Tetradon   |
| 38 | 28.8 | 22.3 | 350  | 13 | A2693856 | A2693856 ASF-11DB2  |
| 39 | 28.8 | 22.3 | 469  | 10 | A1065277 | A1065277 TENU2164   |
| 40 | 28.8 | 22.3 | 772  | 10 | A1316149 | A1316149 vJ25f01.Y  |
| 41 | 28.8 | 22.3 | 949  | 10 | AL531914 | AL531914 AL531914   |
| 42 | 28.6 | 22.2 | 886  | 11 | BF339003 | BF339003 602036082  |
| 43 | 28.6 | 22.2 | 936  | 10 | BE615543 | BE615543 601278885  |
| 44 | 28.4 | 22.0 | 510  | 10 | AW231767 | AW231767 687024C05  |
| 45 | 28.4 | 22.0 | 515  | 10 | BE510594 | BE510594 946054A06  |

#### ALIGNMENTS

RESULT 1  
BG817831/c 600 bp mRNA EST 22-MAY-2001  
LOCUS  
DEFINITION  
SAS0334 5', mRNA sequence.  
ACCESSION  
BG817831  
VERSION  
BG817831.1 GI:14188811  
KEYWORDS  
SOURCE  
ORGANISM  
Sarcoptes scabiei.  
Sarcoptes scabiei.  
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
Acariformes; Sarcoptiformes; Astigmata; Sarcoptoides; Sarcoptidae;  
Sarcoptes.

REFERENCE  
1 (bases 1 to 600)  
Ljunggren, E.L., Nilsson, D., Naslund, K. and Mattsson, J.G.  
Expressed sequence tag analysis of the parasitic mite Sarcoptes  
scabiei

JOURNAL  
COMMENT  
Unpublished (2001)  
Contact: Mattsson J.G.  
Department of Parasitology (SWEPPAR)  
National Veterinary Institute  
SE-751 89 Uppsala, Sweden  
Tel: +46 18 674120  
Fax: +46 18  
Email: jens.mattsson@sva.se  
Similar to p1r17405 scavenger receptor or cysteine-rich protein  
precursor - sea urchin (Strongyloce nitrolos purpuratus).  
Seq primer: 73 primer  
High quality sequence stop: 600.

#### FEATURES

source  
1..600  
/organism="Sarcoptes scabiei"  
/db\_xref="taxon:52283"  
/clone="SAS0334"  
/clone\_lib="S.scabiei cDNA library"







347 GTTTAGGGCATCTGTGGTGAGGCTTTTGCATGTCTTC 309

| FEATURES              | source   |
|-----------------------|--|
| JOURNAL               | Expressed sequence tag analysis of the parasitic mite Sarcopotes scabiei   |
| COMMENT               | <p>Unpublished (2001)</p> <p>Contact: Mattsson J.G.<br/>Department of Parasitology (SMEPAR)<br/>National Veterinary Institute<br/>SE-751 89 Uppsala, Sweden<br/>Tel: +46 18 674120<br/>Fax: +46 18</p> <p>Email: jens.mattsson@sva.se<br/>Similar to IPIR171405 scavenger receptor cysteine-rich protein precursor - sea urchin (Strongyloides nitroctus purpuratus).<br/>Seq primer: T3 primer<br/>High quality sequence stop: 720.</p>   |
| FEATURES              | Location/Qualifiers  |
| source                | 1..720   |
|                       | /organism="Sarcopotes scabiei"   |
|                       | /db_xref="taxon:52283"   |
|                       | /clone="SAS0067"   |
|                       | /note="The Sarcopotes scabiei mixed lifecycle library was constructed by Jens G Mattsson. cDNAs were synthesized from poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the Uni-ZAP lambda vector (Stratagene). The primary library was amplified on XL1-Blue MRF' cells."   |
| BASE COUNT            | 198 a 184 c 163 g 175 t  |
| ORIGIN                |  |
| Query Match           | 23.4%; Score 30.2; DB 11; Length 720;  |
| Best Local Similarity | 56.6%; Pred. No. 39;   |
| Matches               | 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;   |
| OY                    | 2 gaccttcgccctcgtcagagatctactctgagctactctgacgcaagaagcccgaccagagat 61.  |
| Db                    | 590 GACCATGAGGAGCAGCATCTTTGGGGCGATGAGACGACCAAGAGACCCCAATGAGAAAT 531  |
| OY                    | 62 gtcgcggagctcttcgt 100   |
| Db                    | 530 GTTATGAGGACATCTGT 492  |
| RESULT 11             |  |
| CNS017UC/c            |  |
| LOCUS                 | CNS017UC 1101 bp DNA GSS 26-JUL-1999   |
| DEFINITION            | Drosophila melanogaster genome survey sequence SP6 end of BAC BACN31P23 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.  |
| ACCESSION             | AL108510   |
| VERSION               | AL108510.1 GI:5628814  |
| KEYWORDS              | GSS  |
| SOURCE                |  |
| ORGANISM              | Fruit fly.   |
|                       | Plasmod Drosophila melanogaster  |
|                       | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  |
|                       | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;   |
|                       | Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.   |
| REFERENCE             | 1 (bases 1 to 1101)  |
| AUTHORS               | Genoscope.   |
| TITLE                 | Direct Submission  |
| JOURNAL               | Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr   |
| COMMENT               | <p>- Web : www.genoscope.cns.fr)</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBlotBAC11.</p> |
| FEATURES              | Location/Qualifiers  |
| source                | 1..1101  |

```

/organism="Drosophila melanogaster"
/plasmid="pbeloBAC11"
/db_xref="taxon:7227"
/clone_1Id="DrosBAC"
/clone="BACN37P23"
/note="end : SP6"

```

|    |  |                  |                |                     |         |
|----|--|------------------|----------------|---------------------|---------|
|    | Query Match  | 23.4%            | Score 30.2;    | DB 13; Length 1101; |         |
|    | Best Local Similarity  | 53.7%;           | Pred. No. 41;  |                     |         |
|    | Matches 36;  | Conservative 11; | Mismatches 20; | Indels 0;           | Gaps 0. |
| OY | 32 attactgacgaacaagaccgcaggcagaatgtctggcgccatttgtagtgcrtga       | 91               |                |                     |         |
|    | : : - :                 :             :                          |                  |                |                     |         |
| D8 | 1100 ATATCTGGSGSKAMNGSCCAACGCCAAMYTTTGGGKKRYTTTTKTYTGCGITTTTKTGT | 1041             |                |                     |         |
| OY | 92 cgttgttc 98   |                  |                |                     |         |
|    | :   :  |                  |                |                     |         |
| D8 | 1040 GGBRGKG 1034  |                  |                |                     |         |

|            |   |
|------------|---|
| RESULT     | 12  |
| BF841751/c |   |
| LOCUS      | BF841751  |
| DEFINITION | 187 bp mRNA   |
| ACCESSION  | RC2-RT1079-051200-015-e02 HT1079 Homo sapiens CDNA, mRNA sequence |
| VERSION    | BF841751  |
| KEYWORDS   | BF841751.1 GI:12195035  |
| SOURCE     | EST.  |
| ORGANISM   | human.  |
|            | Homo sapiens  |
|            | 13-JAN-2001   |

| REFERENCE          | AUTHORS  | TITLE  |
|--------------------|--|--|
| 1 (bases 1 to 187) | Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bata, G. S., Simpson, D. H., Brindman, A. A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J. | Shotgun sequencing of the human transcriptome with ORF expressed sequence tags |
| JOURNAL<br>MEDLINE |  | Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)                         |
| COMMENT            |  | Contact: Simpson A.J.G.  |

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scisearch/gethtml2.pl?cl=RC2&t2=RC2-HT1079051200-015-e02&t3=2000-12-05&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 187.  
Location/Qualifiers  
1..187

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT1079"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI: A mini-library was made by cloning products
derived from ORSMP5 PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription o
lissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

|            |      |      |      |      |
|------------|------|------|------|------|
| BASE COUNT | 50 a | 59 c | 45 g | 33 t |
| ORIGIN     |      |      |      |      |

[illegible]

|            |   |
|------------|---|
| RESULT     | 13  |
| CNS04GSE   |   |
| LOCUS      | CNS04GSE 903 bp DNA GSS 21-MAY-2000   |
| DEFINITION | Tetradon nigroviridis genome survey sequence T7 end of clone<br>109E21 of library G from Tetradon nigroviridis, genomic survey<br>sequence. |
| ACCESSION  | AL290039  |
| VERSION    | AL290039.1 GI:8028619   |
| KEYWORDS   | GSS; genome survey sequence.  |
| SOURCE     | Tetradon nigroviridis.  |
| ORGANISM   | Tetradon nigroviridis   |

| REFERENCE<br>AUTHORS   | TITLE   | JOURNAL<br>REFERENCE<br>AUTHORS  |
|--|---|--|
| 1 (bases 1 to 903)<br>Roest-Crolius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, N., Bernot, A. and Weissenbach, J. | Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i> | Roest-Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., 2 (bases 1 to 903) |
| Unpublished  |   |  |

|           |  |
|-----------|--|
| TITLE     | Human gene number estimate provided by genome wide analysis using  |
| JOURNAL   | Tetradon nigroviridis DNA sequence   |
| REFERENCE | Unpublished  |
| AUTHORS   | 3 (bases 1 to 903)   |
| TITLE     | Genoscope.   |
| JOURNAL   | Direct Submission  |
| COMMENT   | Submitted (12-APR-2000) to the EMBL/GenBank/DBD databases<br>This sequence is a single read and was generated as part of a large |

| FEATURES   | SOURCE | Location/Qualifiers                                     |
|------------|--------|---|
|            |        | 1..903  |
|            |        | /organism="Tetradodon nigroviridis"                     |
|            |        | /db_xref="taxon:99883"                                  |
|            |        | /clone="109E21"   |
|            |        | /clone_11b="G"  |
|            |        | /note="Genoscope sequence ID : COBG109C111P1-end : T77" |
| BASE COUNT | 231 a  | 195 c 215 g 230 t 32 others                             |
| ORIGIN     |        |   |

|    | Query Match | Similarity   | 25.2%        | Score | 29.6       | DB | 13     | Length | 903  |
|----|-------------|--|--------------|-------|------------|----|--------|--------|------|
| Db | 462         | TTATCTTTTGGTAACTTCCTCTCCACAGTGAATGGAAAAGCCACGGCTGAGCTGCAC    | 521          |       |            |    |        |        |      |
|    | Matches     | 50   | Conservative | 5     | Mismatches | 39 | Indels | 0      | Gaps |
| Qy | 6           | tctcgtcgtagcgaattacttcgcagcattacacgcagacaagaccccgccgagatgctc | 65           |       |            |    |        |        |      |

QY 66 ggggtctttgttggtgctgtgacgtgttt 99  
 Db 522 TGGGTGTTAGGTGTGTATGATGATKTCGT 555

RESULT 14  
 LOCUS H76502 485 bp mRNA EST 05-JAN-1998  
 DEFINITION 18207 Lambda-PRL2 Arabidopsis thaliana cDNA clone 196N13T7, mRNA sequence.  
 ACCESSION H76502  
 VERSION H76502.1 GI:1053753  
 KEYWORDS EST.  
 SOURCE thale cress.  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 485)  
 Newman,T., deBruin,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomasow,M., Metzel,E. and Somerville,C.  
 Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones plant Physiol. 106, 1241-1255 (1994)  
 95148729

JOURNAL MEDLINE  
 COMMENT Contact: Thomas Newman  
 MSU-DOE Plant Research Laboratory  
 Michigan State University  
 MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI  
 Tel: 517-353-0854  
 Fax: 517-353-9168  
 Email: 22333tc@idm.cl.msu.edu  
 Seq primer: 5\_prime.  
 Location/Qualifiers  
 1..485  
 /organism="Arabidopsis thaliana"  
 /strain="var Columbia"  
 /db\_xref="taxon:3702"  
 /clone="196N13T7"  
 /note="Vector: lambda zip-lox. Site\_1: Sal; Site\_2: Not; lambda PRL2 is a cDNA library derived from equal 7 quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda zip-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."

BASE COUNT 130 a 99 c 119 g 118 t 19 others  
 ORIGIN

Query Match 22.8%; Score 29.4; DB 11; Length 485;  
 Best Local Similarity 51.6%; Pred. No. 66;  
 Matches 48; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 26 tcggagcttctgacgaacagccgacgagatgctgtcttctgtgtgtgt 85  
 Db 385 TTTNAGATTAGGNGNGNACCCGNGGAGACANTGCTGTTTGTGTGT 444

QY 86 ctgtgacgtgtgtccaacgtatattccga 118  
 Db 445 CTANNCCTGTTCCTCAATAAATTTCNCGA 477

RESULT 15  
 B1114606 1353 bp mRNA EST 26-JUN-2001  
 LOCUS B1114606 602861984F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:5021228 5',

ACCESSION B1114606 mRNA sequence.  
 VERSION B1114606.1 GI:14565507  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1353)  
 NIH-MGC <http://mgc.ncl.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LICM1839 row: 9 column: 21  
 High quality sequence stop: 71.  
 Location/Qualifiers  
 1..1353  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5021228"  
 /clone\_1ib="NIH\_MGC\_17"  
 /tissue\_type="rhodomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pORF7; Site\_1: EcoRI; Site\_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(5). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life technologies)."

BASE COUNT 320 a 400 c 299 g 333 t 1 others  
 ORIGIN

Query Match 22.8%; Score 29.4; DB 11; Length 1353;  
 Best Local Similarity 52.0%; Pred. No. 74;  
 Matches 66; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 3 acctctctctgtagcattctgcagcattctgcagcacaagaccgacgagatg 62  
 Db 328 ATCTGCGCAGACCGCGCAGTGTGCGCCACCGGAGACAAAGAGCGCGGAGACA 387

QY 63 gtcggggtcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 122  
 Db 388 GTCCGGAAGTATATGACTGCTGTGCTCATCGAGCACTAACCTATACGTAGGCGACACA 447

QY 123 ttcagcg 129  
 Db 448 CAGAGCG 454

Search completed: March 28, 2002, 14:42:27  
 Job time: 7581 sec

